Sortase-Mediated Labeling of M13 Bacteriophage and the Formation of Multi-Phage Structures

Advisor: Professor Angela Belcher Gaelen Thomas Hess

Sortase-mediated labeling of M13 bacteriophage and the formation of multi-phage structures

Abstract

M13 filamentous bacteriophage has been used as a biotemplate for the nucleation of materials. Phage is an ideal and diverse scaffold with its large aspect ratio and ability to display biomolecules to bind a range of targets. To form more complex patterned materials, interactions between the phage must be specific and reliable. We develop a phage labeling method using sortase enzymes to create multi-phage nanostructures.

We exploit two sortases and functionalize the N-termini of the pIII, pIX, and pVIII proteins with small and large moieties. For the pVIII, we show a 100 fold improvement in display of GFP molecules on the phage surface. Taking advantage of orthogonal sortases, we simultaneously label two capsid proteins on a single phage particle. Using these N-terminal labeling techniques, we demonstrate fluorescent staining of cells and construct a lampbrush phage structure linking the pIII of one phage to the pVIII of another using a biotin-streptavidin linkage.

To further expand our labeling repertoire, C-terminal sortase labeling of phage was pursued. To achieve this goal, we transfer a loop structure from cholera toxin to pIII and label it with a fluorophore and a multi-domain protein. With this architecture, we form end-to-end dimers using sortase to conjugate the loop structure to phage containing the nucleophile motif. Lastly, we investigate DNA hybridization as a method for crosslinking phage. Using sortase, we label the pVIII on two sets of phage: one with ssDNA and the other with a complementary DNA oligonucleotide. We anneal these phages together and observe phage networks that are dispersed by heat and reform upon cooling.

Contents

List of Figures

List of Tables

Acknowledgments

I have been fortunate throughout my time in graduate school to be surrounded by people striving to solve some of the most difficult and meaningful problems. During the process, I have been humbled by the struggles, but this has made the successes more rewarding. My accomplishments have been a collaborative effort of everyone I have encountered, and these acknowledgments cannot do justice to how grateful I am to those who have helped me.

I must thank my advisor Angela Belcher, without whom none of my work would have been possible. She provided me with the freedom to follow my interests and truly let my imagination set the limits of my work. Her ambition to attack important problems has motivated me and taught me to never be discouraged by the magnitude of a project.

The Harvard Biophysics program, especially Prof. Jim Hogle and Michele Pfeffer, were wonderful in guiding me through the process while allowing me to pursue my research at MIT without any hardship. I would also like to thank my committee: Profs. Evelyn Hu, Hidde Ploegh, William Shih, and Tim Springer for their input and encouragement. As a collaborator, Hidde was enthusiastic and supportive making both his time and the resources in his lab available.

In my research, I have worked with numerous talented individuals. Thank you to all the Belcher lab members past and present. In particular, I would like to thank Mark Allen, Chung-Yi Chiang, Georg Fantner, Rana Ghosh, David Gray, Nimrod Heldman, Mo Khalil, Joan Mao, and Hyunjung Yi. Mark was helpful with DLS and TEM data presented in this work, while both Rana and Nimrod were extremely helpful and willing to talk out things from the mundane to the outrageous.

I must also thank my collaborators including Yongdae Shin and Prof. Matt Lang from the Lang lab as well as Ana Avalos, Juanjo Cragnolini, Jadyn Damon, Stephanie Dougan, Carla Guimaraes, Max Popp, Eric Spooner, and Kim Swee from the Ploegh lab. Carla Guimiaraes, in particular, deserves special recognition. She is a wonderful collaborator and friend. She taught me more biochemistry (and Portuguese) than I could have expected and she always had time in her busy schedule to talk through an idea. Her drive and enthusiasm were infectious and she made the research process a joy. To Carla, thank you for the help, the encouragement, the advice, the recipes, the movie recommendations, the stories, the laughs, and the smiles. Obrigado, eu não esquecerei tua ajuda.

Outside of the lab, I have a great group of friends who always reminded me to have fun and enjoy life outside of the lab. Vivek Shenoy with our sci-fi B movies, Josh Dezube with his detailed spreadsheets, and Hans Cutiongco with our jam sessions and concerts have always been there for me to help escape work. My friends have been part of memorable vacations including a road trip from New York to Palo Alto, and a "Footy Pilgrimmage" through England and Spain. To all my friends, thank you for being ever present to provide distraction from work or to listen to me talk about it.

My acknowledgments would not be complete without thanking my family who have been integral to all the success I have achieved in my life. My mother would drive me everywhere as a child to make sure I did not miss any opportunities. She stressed education and instilled a strong work ethic for which I am forever grateful. Although she is no longer with us, I know she would be prouder than anyone of my accomplishment. To my dad, my sister, Emily, and brother-in-law, Scott, thank you for lending a kind ear during the tough times.

To everyone with whom I have interacted in the past six years: thank you.

Gaelen Hess

Chapter 1

Background and Introduction

1.1 Introduction

Nature has developed a range of systems to address an array of functions such as light harvesting in plants or the formation of silk fibers in spiders. These systems have been honed over time and provide insight for engineering design in the recapitulation of these functions. From this inspiration, biomimetic approaches have been pursued such as those in the study of the mollusk shell. The combination of the strength, stiffness, and toughness from the nacre structure in mollusk shells has been used as guides for new materials [1]. This precisely ordered hybrid composite consisting of an organic matrix and $CaCO_3$ generates the ability for the shell to dissipate and absorb energy effectively before failure [2]. Researchers have searched for assembly and fabrication methods to mimic this structure and generate its energy dissipation properties [3].

In addition to these biomimetic approaches, biological systems can be engineered to introduce new functionalities or harness an innate one. The self-assembly of biomolecules such as peptides [4], viruses [5–8], and DNA [9, 10] have been investigated for the fabrication of nanoscale structures. DNA origami, which uses DNA as a structural material, is an emerging field of research in nanomaterials and devices [11, 12]. In this technology, connections are made using base pair matching to create complex DNA patterns and structures (Fig. 1.1) [13]. These DNA structures have nanometer features and can serve as chemical and biological sensors programmed to change conformation in response to stimuli such as pH [14]. The function and

complexity of these DNA structures are growing as seen in the creation of a DNA nanorobot [15]. This robot consisted of a DNA scaffold carrying a molecular payload and DNA switches that unlocked the structure in the presence of specific cells to initiate delivery of cargo illustrating how DNA can function as a structural and functional material.

In addition to scaffolds, biological systems supply enzymatic tools commonplace to biological research. *Hind*II was the first characterized DNA restriction enzyme, purified from *Hemophilius influenzae* bacteria, and was found to cut a specific 6bp double-stranded DNA sequence [16]. Since that time, numerous restriction enzymes have been characterized and are standard tools used for molecular cloning [17]. At the protein level, the biotin ligase BirA from *Escherichia coli* (*E. coli*) attaches a biotin group to a specific substrate peptide in the presence of ATP [18]. This enzyme is a common experimental tool for the biotinylation of purified or intracellular proteins [19, 20].

In this work, we harness elements from two biological systems: the M13 bacteriophage scaffold and the enzymatic labeling of sortase A enzymes. Merging these systems, we generate multi-phage scaffolds, which can serve as the basis for complex patterned materials. In this chapter, we provide background on both of these biological systems.

1.2 M13 bacteriophage

M13 bacteriophage is part of the Ff filamentous bacteriophage family. This family infects *E. coli* and uses the host to assemble progeny phage without lysing its host. The M13 phage has a cylindrical shape with a length of 880nm and diameter of 6nm (Fig. 1.2a). The bacteriophage capsid is composed of five proteins (pIII, pVI, pVII, pVIII, and pIX) to encapsulate a single stranded genome. At one end, there

are ∼5 copies of pIX and pVII, which form a complex to start the assembly of the viral capsid [21]. At the other end, there are ∼5 copies of pVI and pIII, which are responsible for infection of the bacteria [21]. The body of the phage is composed of 2700 copies of pVIII and represents 98% of the mass of the phage.

The genome of phage can be manipulated with standard molecular biology techniques to display polypeptides as fusions to the capsid proteins. The connection between the phenotype (the displayed capsid proteins) and the phage genome make the virion a valuable screening tool. In 1985, George Smith introduced phage display to screen peptide libraries on the pIII [22]. In the process (Fig. 1.2b), the phage library is incubated with a target. Unbound phage are washed away and the bound phage are eluted and collected. These phage are amplified and the genomes of the phage are sequenced. With the amplified phage, the process is repeated with the enriched phage population to find phage with the strongest binding properties and a consensus sequence. The pIII is the most common capsid protein used in phage display, but peptide libraries have been expressed on other capsid proteins including pVIII [23].

A second display method, called phagemid display, is used to screen libraries of larger moieties such as single chain antibodies [24]. In phagemid display, a library of inducible plasmids that contain a modified version of a phage capsid protein is transformed into bacteria. The plasmids contain an origin of replication and packaging signal to form a single stranded DNA (ssDNA) version, which can be packaged into assembling phage. The transformed bacteria are infected with a helper phage, which generates the rest of the phage capsid proteins. When the infected bacteria are induced, the capsid protein from the helper phage and the modified version from the phagemid are incorporated into assembling capsids creating hybrid virions. Because every copy of the incorporated capsid protein does not need to contain the modification, the phagemid technique can accommodate larger fusions. However, in order to

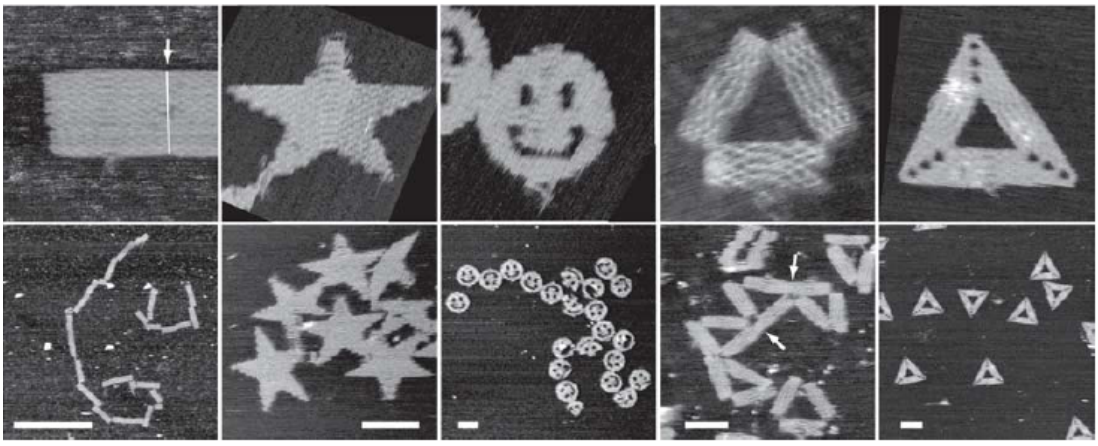


Figure 1.1: **Patterns generated using DNA origami**. Atomic force microscopy images of DNA structures in the shape of a rectangle, star, disc with three holes, triangle with rectangular domains, and triangle with trapezoidal domains are shown (adapted from [13]). The images in the top row are the same size, 165nm by 165nm. The bottom row scale bars are 1μm for the rectangle and 100nm for the other images.

maintain the connection between the phenotype and genotype for screening libraries, the packaging signal of the helper phage genome is impaired such that the single stranded version of the phagemid is given presence to package inside the assembling viral capsids.

Using either peptide or antibody libraries, biomolecules of interest can be identified that bind a desired target such as proteins, cell lines, and inorganics [25–27]. For inorganics, the surface area to volume ratio of M13 bacteriophage allows for efficient nucleation of nanoparticles in a small volume but still exposing a large surface area. Using a pVIII peptide library, an 8-mer amino acid sequence was identified that bound gold nanoparticles. Phage displaying this gold binding peptide sequence was used to form gold nanowires [28]. In addition to material applications, M13 bacteriophage has been used as a scaffold in biomedical applications such as molecular imaging [29, 30] and gene and drug delivery [31, 32].

Although there are limitations to the size of the fusions made with the capsid proteins [33, 34], the phage can serve as a scaffold for the display of rationally designed peptides. Four glutamic acid residues, which are negatively charged at neutral pH, were added onto the N-terminus of pVIII to create a construct that nucleates metallic materials through electrostatic interactions [35]. This phage has been used to nucleate inorganics such as iron phosphate, silver, and cobalt oxide (Fig. 1.3) [35–37]. Phage modified using a selection based or a rational design approach has been used as a material scaffold to form battery, solar cell, and photocatalyst devices [35, 38, 39].

In addition to genetic approaches of display, phage has been used as a scaffold for chemical modification [40]. These strategies include the incorporation of reactive unnatural amino acids such as selenocysteine [41] or azidohomoalanine [42]. Natural amino acids such as cysteines, arginines, lysines, and tryptophans may function as targets for chemical modification as well [40]. These chemistries often lack specificity to a single capsid protein, which limits their utility in making phage to phage

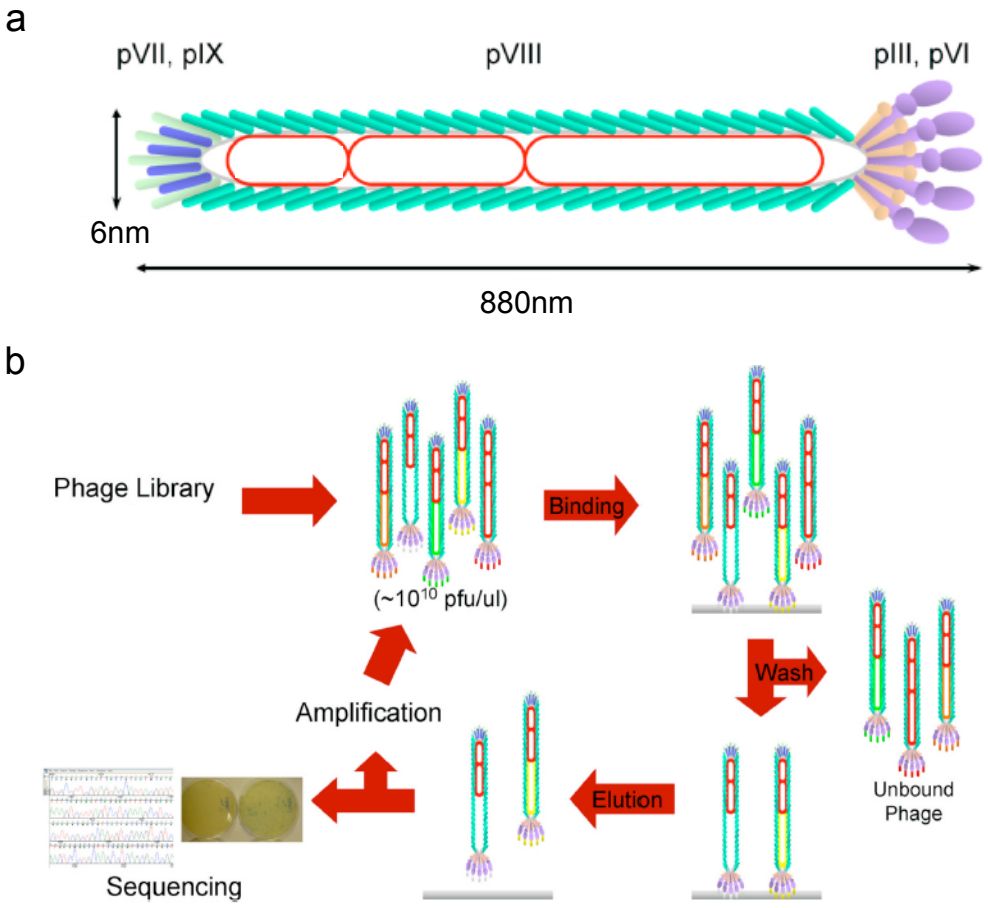


Figure 1.2: **M13 bacteriophage structure and phage display**. M13 bacteriophage is cylindrical and consists of five capsid proteins (pIII, pVI, pVII, pVIII, and pIX) encasing a single stranded genome (a). Schematic representation of the phage display process where peptides are identified from libraries that bind a desired target (b).

interactions between specific capsid proteins, which is necessary for controlling the orientation of phage.

1.3 Sortase enzymes

Sortase A enzymes are bacterial proteins that are used to affix peptides and proteins to the cell wall [43]. These sortase enzymes recognize a peptide motif which acts as a substrate for the enzyme forming an intermediate that is resolved by a nucelophile motif. In this work, two sortases are discussed: Sortase A from *Staphylococcus aureus* ($\mathrm{SrtA_{aureus}}$) and *Streptococcus pyogenes* ($\mathrm{SrtA_{pyogenes}}$). These sortases recognize an LPXTG and LPXTA motif, respectively [44, 45]. The enzymes cleave after the threonine forming a thioester linkage between the carboxyl end of the threonine and the sulfhydryl group on the sortase [45] (Fig. 1.4). This acyl-enzyme intermediate is resolved by the amino group on the end of an oligoglycine or oligoalanine motif, respectively.

Sortase enzymes have been developed as a tool for protein labeling [46]. The sortase chemistry offers advantages over other labeling chemistries. For labeling proteins, it is specific as the motifs for labeling are not common. Different sortases can orthogonally label a mixture of proteins or multiple locations on a single protein [47]. With this labeling method, proteins and cells have been conjugated with a range of substrates including biotins, fluorophores, lipids, and proteins [48–50].

Recently, this labeling chemistry has been used to study the trafficking of cholera toxin [48]. The catalytic subunit of the toxin was engineered to include the substrate motif and labeled with a biotinylated reporter peptide that is glycosylated upon reaching the endoplasmic reticulum (ER). The biotin group provides a mechanism for detection of the subunit by immunoblot. Using this construct, the trafficking of cholera toxin to the ER was confirmed and quantified. Additionally,

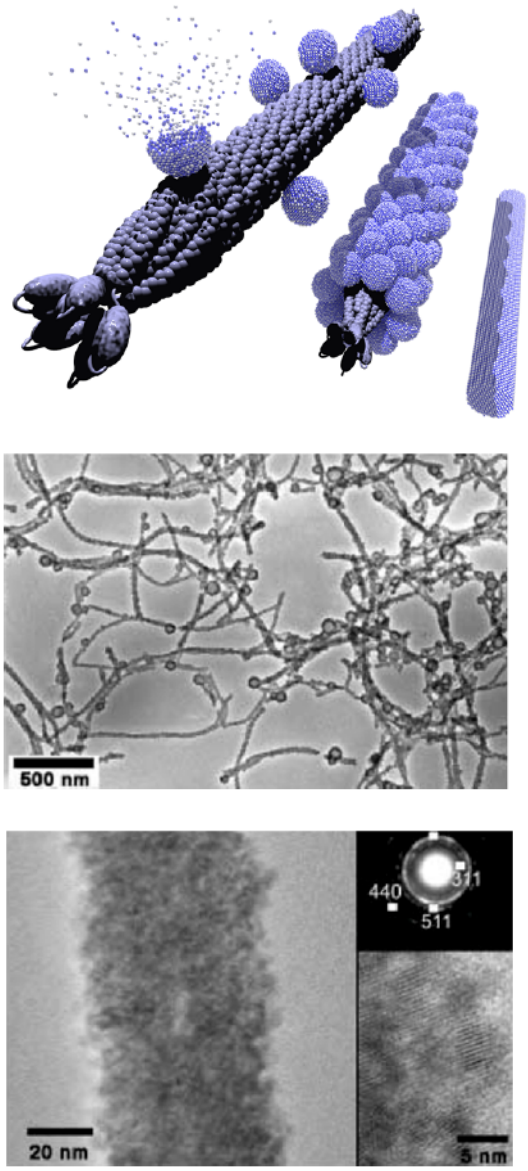


Figure 1.3: **Biotemplation of M13 bacteriophage**. M13 bacteriophage has been used as a biotemplate facilitating the synthesis of nanowires. A visual representation of this process is shown (top panel) and has been demonstrated for Co_3O_4 nanowires for use in lithium ion battery electrodes (bottom panels) [35].

using sortase, the catalytic subunit of diphtheria was connected to cholera toxin to generate a lethal construct that undergoes the trafficking of cholera toxin. Using this construct, Guimaraes *et al.* were able to perform a genetic screen and identify genes critical to the trafficking of cholera toxin.

1.4 Scope of work

Although M13 bacteriophage has been used to construct devices through the process of biotemplating, there is a limit to what can be achieved with current techniques for connecting and ordering phage. To build more complex scaffolds, the interactions between the phage must be reliable and specific.

In this dissertation, we develop the sortase labeling method on the M13 bacteriophage platform and create a variety of connected phage structures. In Chapter 2, we engineer phage to participate in sortase-mediated reactions and install biotinylated peptides and GFP molecules on the pIII, pIX, and pVIII capsid proteins using either $SrtA_{aureus}$ or $SrtA_{pyogenes}$. We quantify the labeling and show a significant improvement over phagemid display of large proteins for the pVIII. Using this conjugation method, we show the ability to fluorescently stain specific cells using phage labeled with a fluorophore and a targeted antibody. Using streptavidin and biotin groups placed on the phage surface by sortase, we generate the lampbrush phage structure by connecting the head of one phage to the body of another.

In Chapter 3, we display the $SrtA_{aureus}$ substrate motif by installing a loop structure from cholera toxin onto pIII. We demonstrate that the loop can be labeled with a fluorophore and a multi-domain protein by $SrtA_{aureus}$. With this architecture and using sortase, we fuse the sortase substrate motif of one phage to the nucleophile motif on the pIII of another phage forming end-to-end structures. In Chapter 4, we investigate the formation of recyclable phage networks. We build such networks

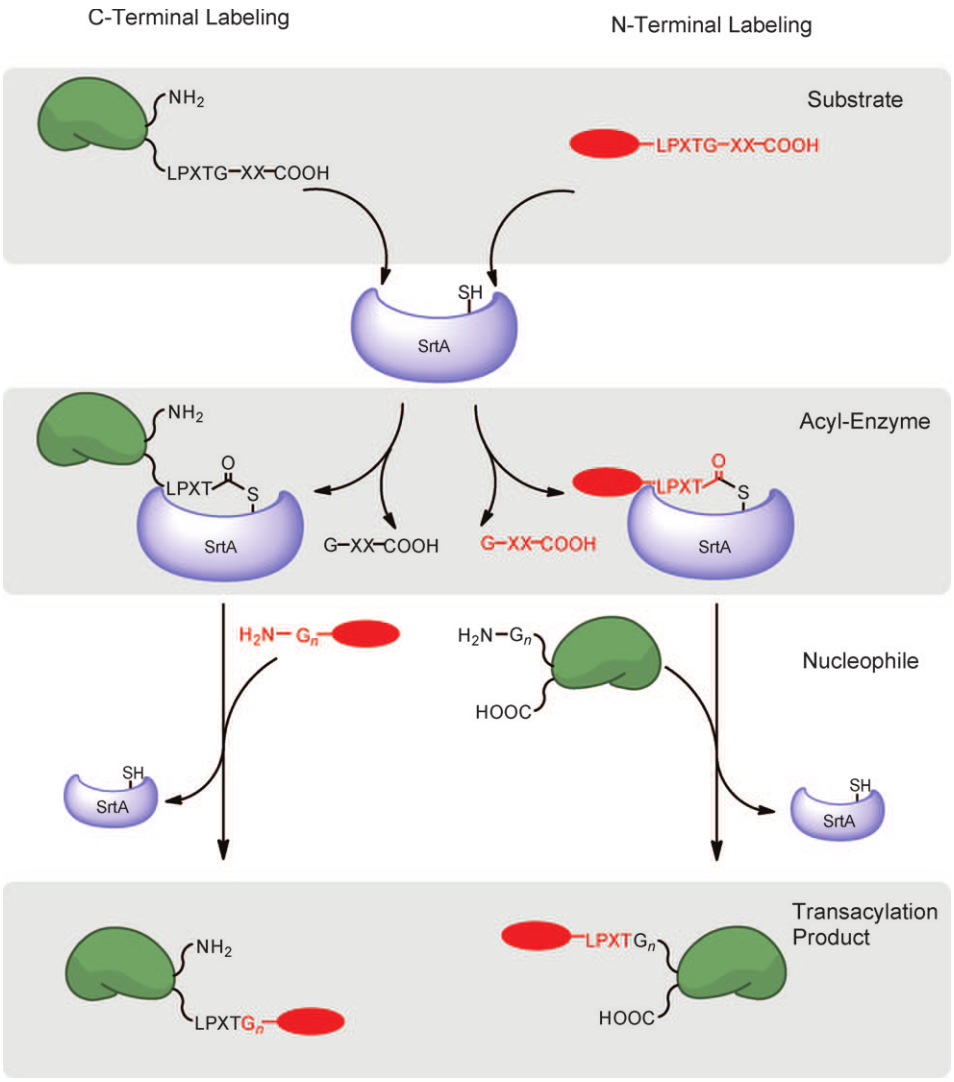


Figure 1.4: **C- and N-terminal sortase-mediated labeling**. C- and N-terminal labeling using sortase A from *Staphylococcus aureus* ($\text{SrtA}_{\text{aureus}}$). This figure is taken from [51].

through the installation of complementary strands of DNA on two phage and anneal the phage together by DNA hybridization. Once connected, we demonstrate that the networks are dispersed by heating and reform upon cooling.

Chapter 2

Sortase-mediated labeling of M13 bacteriophage

2.1 Introduction

This chapter has been adapted from [52].

M13 bacteriophage has a cylindrical shape with a length of 880nm and a diameter of 6nm. It encapsulates a single-strand genome that encodes five different capsid proteins (Fig. 2.1). The body of the phage is composed of 2700 copies of pVIII, the major capsid protein. At one end of the virus, there are ∼5 copies of both pIII and pVI proteins, and at the other end there are ∼5 copies of both pVII and pIX proteins [53].

The capsid proteins of M13 bacteriophage have been used to express combinatorial peptide libraries or protein variants (ranging from single domains to antibodies) to screen for target ligands in a process known as phage display [54]. This technique has enabled not only identification of peptides with affinity for biological targets such as proteins, cells, and tissues [55–58], but also allowed the identification of biomolecules that bind inorganics [27, 59]. These molecules, when expressed on the M13 capsid proteins, can serve as scaffolds for nanowires, structures, and devices [5, 6, 35, 38, 39].

Functionalization of a virion capsid such as M13 is currently accomplished using chemical and/or genetic approaches [40, 60]. However both strategies have limitations. Chemical conjugations are convenient and versatile, but they label motifs

found on multiple M13 capsid proteins and oftentimes require non-physiological pH and reducing conditions that compromise the activity of the molecule that is being attached or of the moieties already displayed on other capsid proteins [40]. Genetic engineering of phage allows the encoded protein/peptide to be displayed precisely [36, 38], but it has intrinsic restrictions. Two classes of vectors are available for genetic phage display: phagemid and phage. A phagemid allows expression of large fusions with any of the five M13 phage capsid proteins, but these fusions are incorporated at low efficiency [61–65]. In a phage vector, the M13 bacteriophage genome is modified directly. As a result, every copy of the recombinant capsid protein incorporated into the virus displays the modified protein. However, this strategy does not support display of large moieties [66–68]. pVIII allows the display of a larger number of recombinant molecules per phage particle, but it also has the strictest size limitation in phage vector display. pVIII peptide libraries are mostly limited to sizes of up to 10 amino acids, as phage with longer insertions rarely assemble [33, 34]. Insertions of 6-20 amino acids onto pVIII are possible using phagemid, but their display is inefficient with less than 25% of the copies of pVIII containing the desired fusion product [64]. Incorporation of proteins is even less efficient on pVIII: a 23kDa protein is displayed, on average, on less than a single copy of the pVIII fusion per phage particle using a phagemid vector [62]. Phage display methods on the pVIII have been able to increase the binding affinity of phage displaying a moiety [67], but the displayed copy number of the moiety has not been determined. Large moieties of at least 23kDa have been genetically fused to all four minor capsid proteins using a phagemid vector [66, 69, 70], but only pIII has been extensively used in the phage vector system [71]. However, viability of the resultant phage fusions does not guarantee that the recombinant peptide/protein of interest displays its native structure. Both the environment where phage assembles and the phage coat protein to which the protein of interest is fused may interfere with proper folding [72]. This is par-

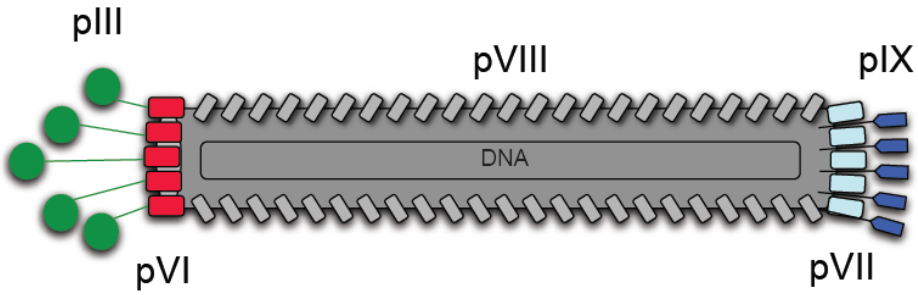


Figure 2.1: **M13 bacteriophage structure**. M13 bacteriophage is composed of five capsid proteins. pVIII is the major capsid protein with ∼2700 copies in each phage particle. The pVII (light blue) and pIX (blue) are located at one end and start the assembly process, while pIII (green) and pVI (red) are at the other end and cap the phage. Note: the image is not to scale.

ticularly critical for enzymes and antibodies, as they might not be functional when incorporated into the phage structure.

To expand the versatility of M13 as a display platform, we devised a strategy based on sortase-mediated chemo-enzymatic reactions to covalently attach a variety of moieties to the N-terminus of pIII, pIX, and pVIII. We obtained a vast improvement over the published data in the copy number of displayed peptides and proteins, particularly on pVIII.

Sortase A enzymes allow modification of proteins by enzymatic ligation with a wide range of functional groups (including biotin, fluorophores, and other proteins) at the C-terminus, N-terminus, or at both termini of the protein of interest [46–50]. The widely used sortase A from *Staphylococcus aureus* ($\mathrm{SrtA_{aureus}}$) recognizes substrates that contain an LPXTG sequence [44, 51, 73], whereas sortase A from *Streptococcus pyogenes* ($\mathrm{SrtA_{pyogenes}}$) recognizes substrates with an LPXTA motif [45, 47]. The sortase enzymes cleave between the threonine and glycine or alanine residue, respectively, to yield a covalent acyl-enzyme intermediate that is resolved by nucleophilic attack of a suitably exposed amine, namely oligoglycine or oligoalanine-containing peptides [45] in the case of $\mathrm{SrtA_{aureus}}$ or $\mathrm{SrtA_{pyogenes}}$, respectively (Fig. 2.2). Using these two orthogonal sortase A enzymes [47], we established conditions to site-specifically attach two different moieties onto two capsid proteins in a single phage particle.

The sortase labeling method has several advantages over genetic and chemical methods. First, the reaction is site-specific, as none of the coat proteins naturally display the required motifs to participate in sortase-mediated reactions. Second, these motifs are small and, therefore, can be easily inserted into the phage genome, maximizing the number of potential attachment sites. Third, a protein to be displayed on phage by means of sortase can be properly folded separate from the assembly of phage. The site-specific nature of the reaction fixes the orientation of the displayed protein. Fourth, the reactions are performed under physiological conditions. Fifth,

sortase reactions afford attachment of a wide range of molecules, including those that cannot be genetically encoded such as fluorophores and biotin.

This work establishes the component parts to build phage structures that have new material and biological applications. We provide two examples: the creation of a new lampbrush structure by fusing different phage particles through pIII/pVIII and a fluorescently labeled phage containing a cell-targeting moiety to stain and to FACS sort cells.

2.2 Results

2.2.1 N-terminal labeling of pIII using $\mathbf{SrtA_{aureus}}$

pIII has been the most extensively explored of the M13 capsid proteins in phage display because of the flexibility and accessibility of its N-terminus [74]. Thus, we introduced five glycines at the N-terminus of pIII (G_5-pIII phage) and used $SrtA_{aureus}$ to covalently attach a K(biotin)-LPETGG peptide (Fig. 2.3). The biotin moiety allowed us to monitor the reaction by immunoblot analysis using streptavidin-HRP. Only when sortase, G_5-pIII phage, and the peptide are incubated together did we detect a 55kDa streptavidin and anti-pIII reactive protein band (Fig. 2.3). The reaction was specific: no other phage proteins were biotinylated. After 3hrs at 37°C, we achieved a yield of 68±9% labeling using 50μM peptide, 50μM $SrtA_{aureus}$, 200nM G_5-pIII phage, and 10mM $CaCl_2$. The efficiency of the reaction was calculated using densitometric analysis of immunoblots where we compared the signal of the biotinylated pIII to biotinylated GFP standards of known concentration. The amount of biotinylated pIII was then divided by the amount of pIII molecules loaded onto the gel, as determined by UV-vis spectrometry. The quantification was repeated for three independent reactions with three samples analyzed for each reaction. The method of quantification is described in further detail in the Methods section.

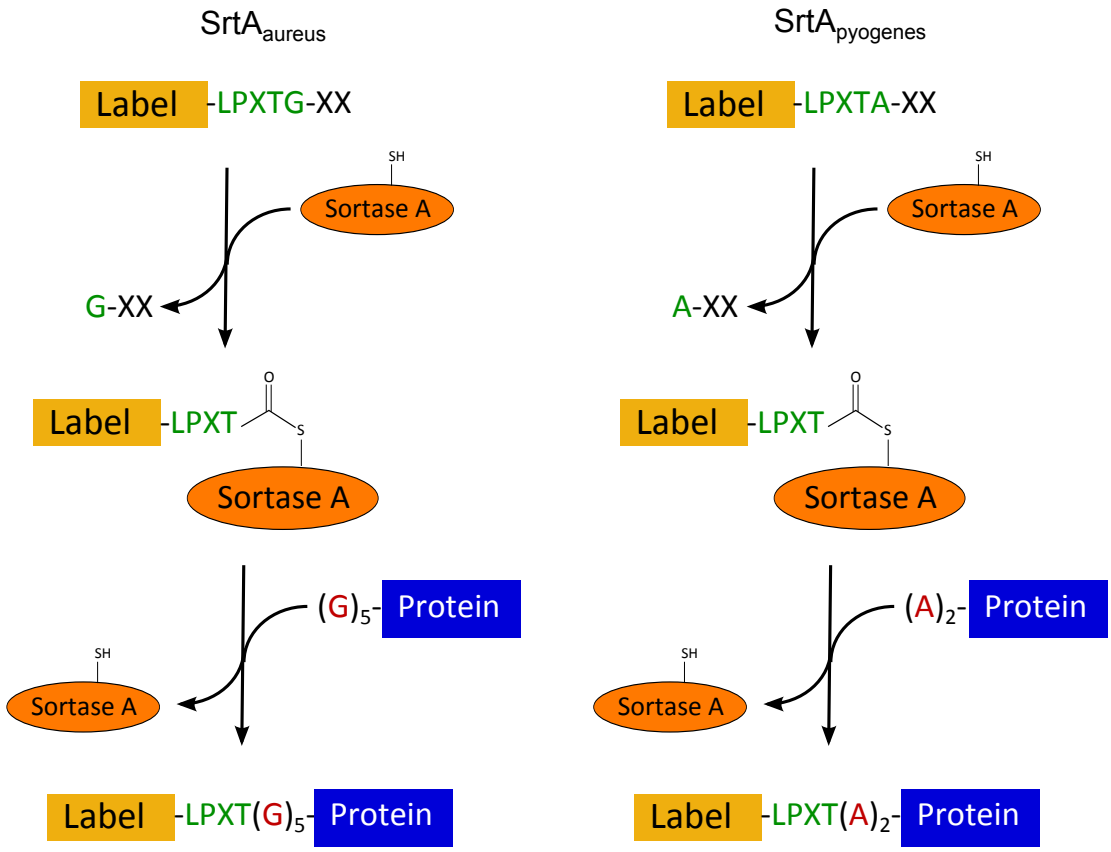


Figure 2.2: **Schematic of sortase-mediated labeling**. Schematic representation of the mechanism of chemo-enzymatic labeling mediated by *Staphylococcus aureus* ($SrtA_{aureus}$-left) or *Streptococcus pyogenes* ($SrtA_{pyogenes}$-right).

To determine whether sortase could be exploited to attach pre-folded proteins onto pIII, we used GFP containing an LPETG motif at its C-terminus as a substrate. The reaction was analyzed by immunoblot using an anti-pIII antibody (Fig. 2.4). Upon completion of the reaction, a mobility shift of pIII to the ∼80kDa region, corresponding to the GFP-pIII fusion product, was detected. The identity of this material was confirmed by mass spectrometry (Fig. 2.4 and Supplementary Fig. A.1). After 3hrs at room temperature, we achieved a yield of 56±2% labeling using 20μM GFP-LPETG, 50μM $SrtA_{aureus}$, 200nM G_5-pIII phage, and 10mM $CaCl_2$. The reaction was quantified by densitometry comparing the signal of pIII-GFP to the signal of the intact pIII input loaded into the reaction.

2.2.2 N-terminal labeling of pIX using $SrtA_{aureus}$

Because the C-terminus of pIX is buried in the phage structure and therefore unavailable for labeling [75], we attempted to label its N-terminus. However, this region of the protein is not as accessible as in pIII and our first attempts at labeling a phage construct displaying five glycines at the N-terminus of pIX using sortase failed (data not shown). To increase accessibility of the five glycines, the N-terminus of pIX was extended with an HA tag, a useful handle for detection, as no pIX-specific antibodies are available. This G_5HA-pIX phage construct was labeled with the K(biotin)-LPETGG peptide and the reactions were analyzed by immunoblot using streptavidin-HRP and an anti-HA antibody. A 5kDa polypeptide, reactive with both streptavidin and anti-HA, was seen only in the complete reaction (Fig. 2.5). We achieved a yield of 73±2% using 50μM peptide, 50μM $SrtA_{aureus}$, 200nM G_5HA-pIX phage, and 10mM $CaCl_2$ upon incubation at 37°C for 3hrs. A similar efficiency was attained when attaching GFP to pIX: 74±1% of pIX was labeled when 20μM GFP-LPETG, 50μM $SrtA_{aureus}$, 200nM G_5HA-pIX phage, and 10mM $CaCl_2$ were incubated for 3hrs at room temperature. A 35kDa anti-HA reactive polypeptide - consistent

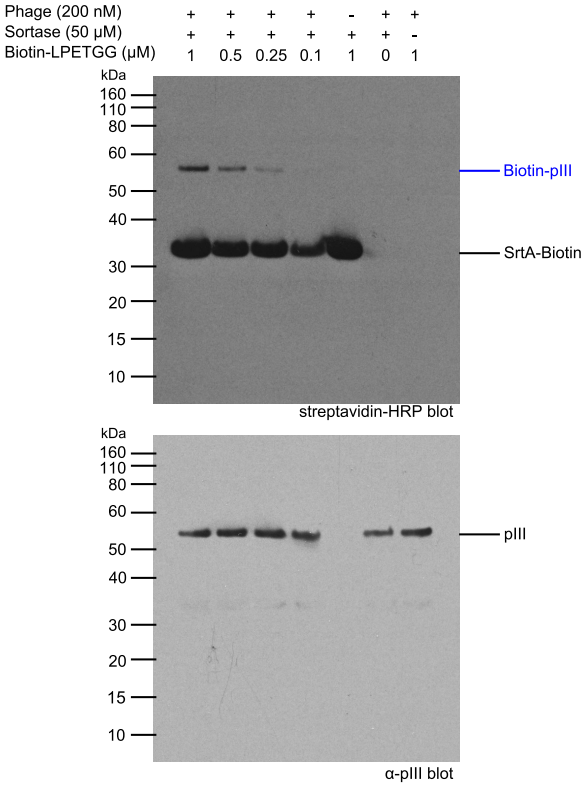


Figure 2.3: **pIII labeling with biotin**. G_5-pIII modified phage (there are five copies of pIII/ phage particle) was incubated with $SrtA_{aureus}$ and K(biotin)-LPETGG peptide for 3hrs at 37°C. The reactions were monitored by SDS-PAGE under reducing conditions followed by immunoblotting using streptavidin-HRP (top panel) or an anti-pIII antibody (bottom panel). The molecular weight markers are shown on the left.

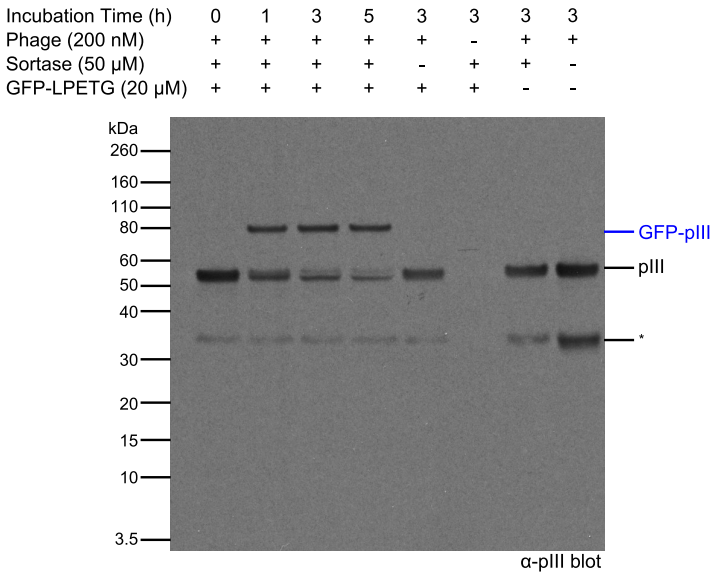
with the molecular mass of the GFP-pIX fusion protein - was detected only in the complete reaction and its identity was confirmed by mass spectrometry (Fig. 2.6 and Supplementary Fig. A.2).

2.2.3 N-terminal labeling of pVIII using $SrtA_{pyogenes}$

In the course of phage biogenesis the N-terminus of pVIII is proteolytically cleaved, resulting in the display of an N-terminal alanine [76]. We took advantage of this feature and exploited $SrtA_{pyogenes}$ to label pVIII. Also, the ability of using two orthogonal sortase enzymes ($SrtA_{pyogenes}$ for pVIII and $SrtA_{aureus}$ for pIII and pIX labeling) would further enable dual labeling of the same phage particle.

To be used as a nucleophile in $SrtA_{pyogenes}$-mediated reactions, pVIII requires display of two N-terminal alanines. Thus, the N-terminus of the mature form of pVIII was modified to AAGGGG (A_2G_4-pVIII phage). The glycines were introduced to extend the N-terminus of pVIII away from the body of the phage, thus improving the accessibility of the Ala-Ala motif for participation in the sortase reaction. Using $SrtA_{pyogenes}$ and a K(biotin)-LPETAA substrate peptide, we showed robust labeling of pVIII based on an immunoblot using streptavidin-HRP (Fig. 2.7). Only when A_2G_4-pVIII phage, $SrtA_{pyogenes}$, and the peptide were mixed together did we detect a biotinylated 10kDa protein, consistent with the size of pVIII. The labeling reaction was site-specific as no other proteins can be detected in the blot. We obtained a yield of 50±3% labeled pVIII when reactions were performed at 37°C for 3hrs with 20μM peptide, 50μM $SrtA_{pyogenes}$, and 8nM A_2G_4-pVIII phage. This translated to 1350±90 biotin molecules on average per phage particle.

Phage assembly limits either the size of the modifications displayed on pVIII to a few residues when using a phage vector, or it limits the number of labels attached to pVIII when using a phagemid vector [64]. In this context, the sortase-labeling strategy is an obvious alternative to overcome such limitations. Using 20μM GFP containing



MVSKGEELFT GVVPILVELD GDVNGHKFSV SGEGEGDATY
GKLTLKFICT TGKLPVPWPT LVTTLTYGVQ CFSRYPDHMK
QHDFFK**SAMP EGYVQERTIF FKDDGNYK**TR **AEVKFEGDTL**
VNRIELK**GID FKEDGNILGH KLEYNYNSHN VYIMADK**QKN
GIKVNFKIRH NIEDGSVQLA DHYQQNTPIG DGPVLLPDNH
YLSTQSALSK DPNEKRDHMV LLEFVTAAGI TLGMDELYK**L**
PETGGGGGSA ETVESCLAKS HTENSFTNVW KDDKTLDRYA
NYEGCLWNAT GVVVCTGDET QCYGTWVPIG LAIPENEGGG
SEGGGSEGGG SEGGGTKPPE YGDTPIPGYT YINPLDGTYP
PGTEQNPANP NPSLEESQPL NTFMFQNNRF RNR**QGALTVY**
TGTVTQGTDP VKTYYQYTPV SSKAMYDAYW NGKFRDCAFH
SGFNEDLFVC EYQGQSSDLP QPPVNAGGGS GGGSGGGSEG
GGSEGGGSEG GGSEGGGSGG GSGSGDFDYE KMANANK**GAM**
TENADENALQ SDAKGKLDSV ATDYGAAIDG FIGDVSGLAN
GNGATGDFAG SNSQMAQVGD GDNSPLMNNF RQYLPSLPQS
VECRPFVFGA GKPYEFSIDC DKINLFRGVF AFLLYVATFM
YVFSTFANIL RNKES

Figure 2.4: **pIII labeling with GFP**. G_5-pIII modified phage (there are five copies of pIII/ phage particle) was incubated with $SrtA_{aureus}$ and GFP-LPETG for 3hrs at room temperature. The reactions were monitored by SDS-PAGE under reducing conditions followed by immunoblotting with an anti-pIII antibody. The molecular weight markers are shown on the left. The unidentified anti-pIII reactive protein (*) is most probably a proteolytic fragment of pIII. The identity of the GFP-pIII fusion product was determined by mass spectrometry. The amino acid sequences of pIII and GFP are shown in blue and green, respectively. The peptides identified are highlighted in bold. The tryptic peptide comprising the GFP C-terminus, followed by the $SrtA_{aureus}$ cleavage site, fused to the N-terminal glycines of pIII is shown in red.

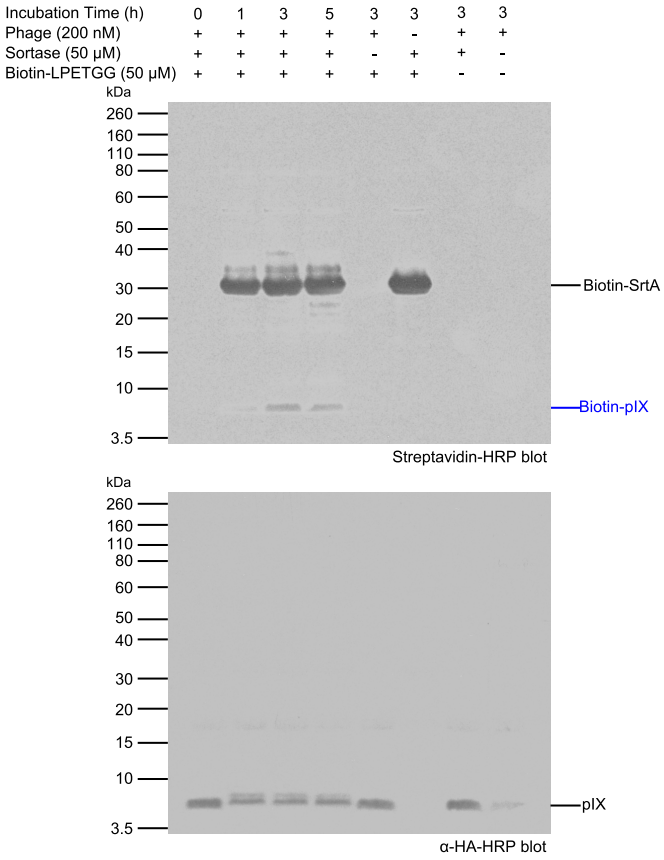
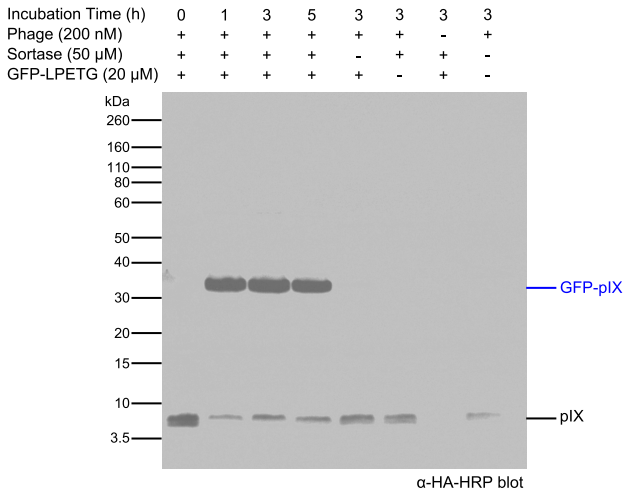


Figure 2.5: **pIX labeling with biotin**. G_5HA-pIX modified phage (there are five copies of pIX /phage particle) was incubated with $SrtA_{aureus}$ and K(biotin)-LPETGG peptide at 37°C for the times indicated. The reactions were monitored by SDS-PAGE under reducing conditions followed by immunoblotting using streptavidin-HRP (top panel) or an anti-HA antibody (bottom panel). The molecular weight markers are shown on the left.



MVSKGEELFT GVVPILVELD GDVNGHKFSV
SGEGEGDATY GKLTLKFICT TGKLPVPWPT
LVTTLTYGVQ CFSRYPDHMK QH**DFFKSAMP**
EGYVQERTIF FKDDGNYKTR AEVKFEG**DTL**
VNRIELKGID FKEDGNILGH KLEYNYNSHN
VYIMADKQKN GIKVNFKIRH NIEDGSVQLA
DHYQQNTPIG **DGPVLLPDNH YLSTQSALSK**
DPNEKRDHMV LLEFVTAAGI TLGM**DELYKL**
PETGGGGGYP YDVPDYAQGG QGVDMSVLVY
SFASFVLGWC LRSGITYFTR LMETSS

Figure 2.6: **pIX labeling with GFP**. G_5HA-pIX modified phage (there are five copies of pIX /phage particle) was incubated with $SrtA_{aureus}$ and GFP-LPETG at room temperature for the times indicated. The reactions were monitored by SDS-PAGE under reducing conditions followed by immunoblotting using an anti-HA antibody. The molecular weight markers are shown on the left. The identity of the GFP-pIX fusion product was determined by mass spectrometry. The amino acid sequences of pIX and GFP are shown in blue and green, respectively. The peptides identified are highlighted in bold. The AspN digestion-resultant peptide comprising the GFP C-terminus, followed by the $SrtA_{aureus}$ cleavage site, fused to the N-terminal glycines of pIX is shown in red.

a LPETA motif at its C-terminus, 50μM $SrtA_{pyogenes}$, and 8nM A_2G_4-pVIII phage, we were able to attach 91±20 GFP molecules on average per phage particle upon incubation at 37°C for 3hrs (Fig. 2.8). The identity of the 35kDa anti-GFP reactive protein, consistent with the size of a GFP-pVIII fusion protein, was confirmed by mass spectrometry (Fig. 2.8 and Supplementary Fig. A.3). As estimated by nearest neighbor packing (described in Appendix A), a single virion can accommodate 385 copies of GFP on its surface. Thus, using the sortase-mediated reaction, we obtained a yield of ∼25% of estimated maximum packing.

2.2.4 Building end-to-body phage structures

The ability to site-specifically label the M13 capsid proteins provides the opportunity to create novel multi-phage structures, which may provide scaffolds for new materials and devices. One such structure (Fig. 2.9) relies on tight binding of the ends of several phage particles (via either pIII or pIX) to the body of another single phage (onto pVIII). However, direct covalent attachment between two phage proteins is not possible using sortase as we were unable to label the C-terminus of pIII, pIX, or pVIII (data not shown). This issue was solved by attaching streptavidin to pIII, biotin to pVIII, and then mixing the two preparations.

Streptavidin, modified to contain a C-terminal LPETG motif in each of its monomers, was attached to the G_5-pIII phage using $SrtA_{aureus}$. The samples were boiled, loaded onto an SDS-PAGE gel, and analyzed by immunoblot using an anti-pIII antibody. A 90kDa polypeptide, consistent with the size of pIII fused to a streptavidin monomer, was seen only when all the reaction components were mixed together (Fig. 2.10). The streptavidin-pIII phage was purified from sortase and free streptavidin by PEG/NaCl precipitation. Dynamic light scattering (DLS) was performed in order to monitor dispersity and aggregation. The normalized autocorrelation function (ACF) of streptavidin-pIII phage showed an exponential decay consistent with monodisperse

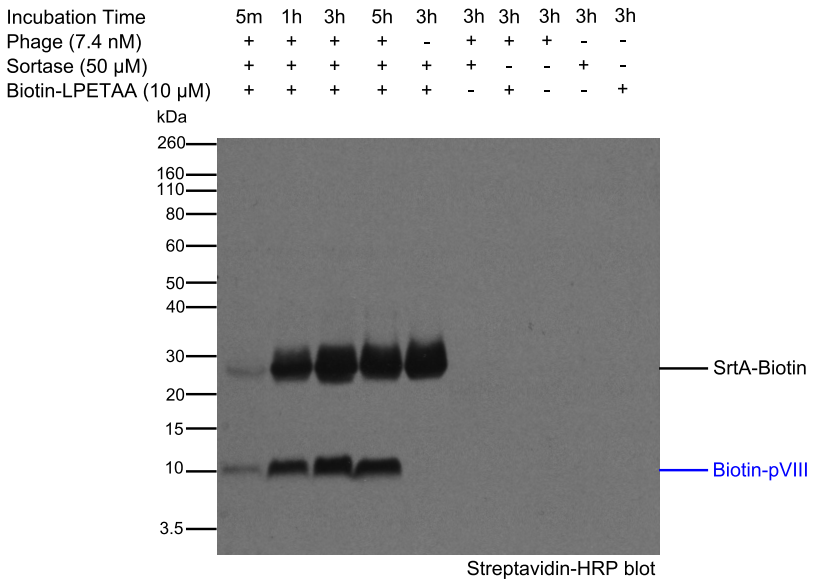


Figure 2.7: **pVIII labeling with biotin**. A_2G_4-pVIII modified phage (there are 2700 copies of pVIII/ phage particle) was incubated with $SrtA_{pyogenes}$ and K(biotin)-LPETAA peptide at 37°C for the times indicated in the figure. The reactions were monitored by SDS-PAGE under reducing conditions followed by immunoblotting using streptavidin-HRP. The molecular weight markers are shown on the left.

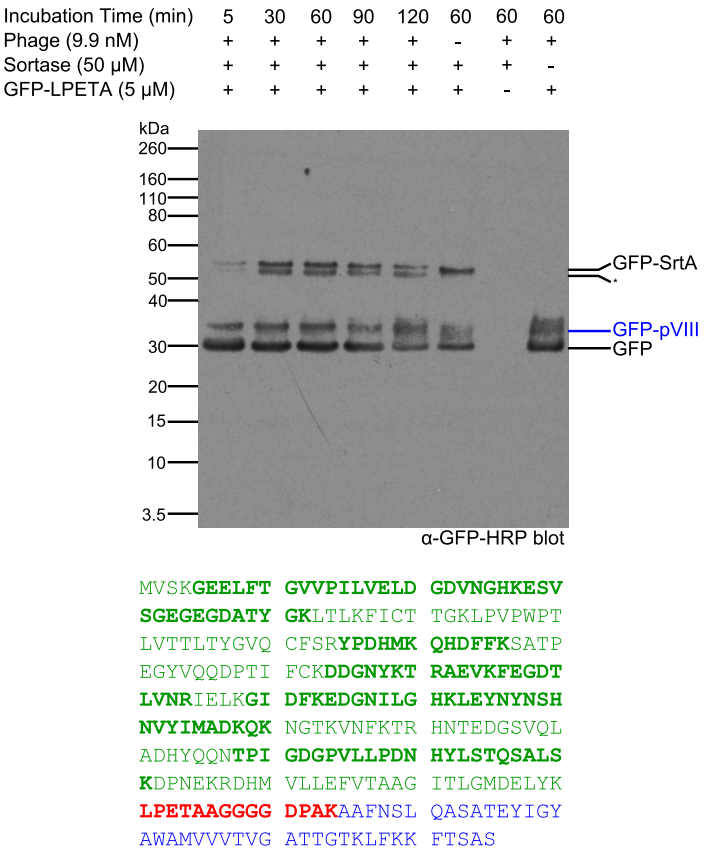


Figure 2.8: **pVIII labeling with GFP**. A_2G_4-pVIII modified phage (there are 2700 copies of pVIII/ phage particle) was incubated with $SrtA_{pyogenes}$ and GFP-LPETA at 37°C for the times indicated in the figure. The reactions were monitored by SDS-PAGE under reducing conditions followed by immunoblotting an anti-GFP antibody. The molecular weight markers are shown on the left. The unidentified anti-GFP reactive protein (*) is attributed to proteolyzed GFP forming an intermediate with $SrtA_{pyogenes}$. The identity of the GFP-pVIII fusion product was determined by mass spectrometry. The amino acid sequences of pVIII and GFP are shown in blue and green, respectively. The peptides identified are highlighted in bold. The tryptic peptide comprising the GFP C-terminus, followed by the $SrtA_{pyogenes}$ cleavage site, fused to the N-terminal alanines of pVIII is shown in red.

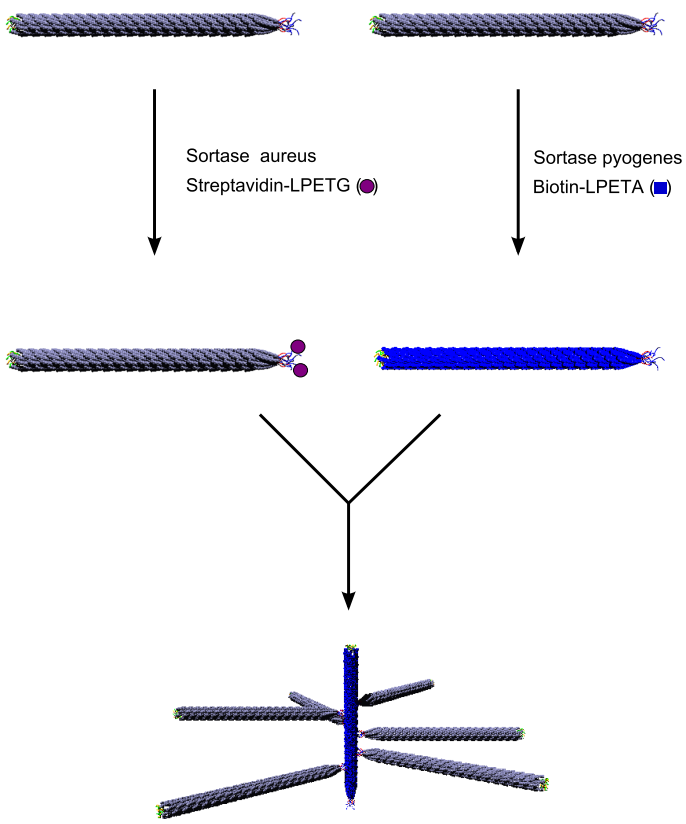


Figure 2.9: **Schematic of lampbrush phage structure**. Schematic representation of the strategy used to build a lampbrush structure. Upon labeling of the N-terminus of pIII with streptavidin and of the N-terminus of pVIII with biotin using sortase-mediated reactions, the phage were mixed (see Methods section for details).

populations (Fig. 2.11a). This was confirmed by atomic force microscopy (AFM) that showed individual virions, indicating that only a single phage particle was attached per streptavidin tetramer (Fig. 2.12). Biotin was conjugated to pVIII using the K(biotin)-LPETAA peptide and $SrtA_{pyogenes}$ as described above. The biotinylated phage was purified by PEG/NaCl precipitation to remove free peptide and the sortase-acyl intermediate. The biotinylated phage was observed as individual phage particles by AFM and the ACF showed an exponential decay, again indicating a monodisperse population (Fig. 2.11a and Fig. 2.12).

The streptavidin-pIII phage and the biotin-pVIII phage were mixed at a 5:1 molar ratio and incubated at room temperature for 15min. Analysis of these samples by DLS showed an increase of the hydrodynamic diameter for the lampbrush phage mixture (700nm) when compared to streptavidin-pIII (516nm) and biotin-pVIII (204nm) phage preparations. When the two types of phage were mixed, the ACF (Fig. 2.11a) shows a rising shoulder at longer relaxation times, indicating a polydisperse population. The longer relaxation times observed in the shoulder represent structures larger than single phage. These larger structures were observed by AFM (Fig. 2.11b). Linkages between the end of one phage and the body of another phage were observed when streptavidin-pIII and biotin-pVIII are mixed. These linkages were not detected when the individual phages were visualized by AFM (Fig. 2.12).

2.2.5 Site-specific labeling of two capsid proteins in the same phage particle

The two orthogonal sortases used to label different capsid proteins offer the possibility to attach different moieties to the body (using $SrtA_{pyogenes}$) and to the end of phage (using $SrtA_{aureus}$) within the same virion. In such a strategy, either pIII or pIX could be labeled with $SrtA_{aureus}$ orthogonally to the pVIII, so as a proof-of-

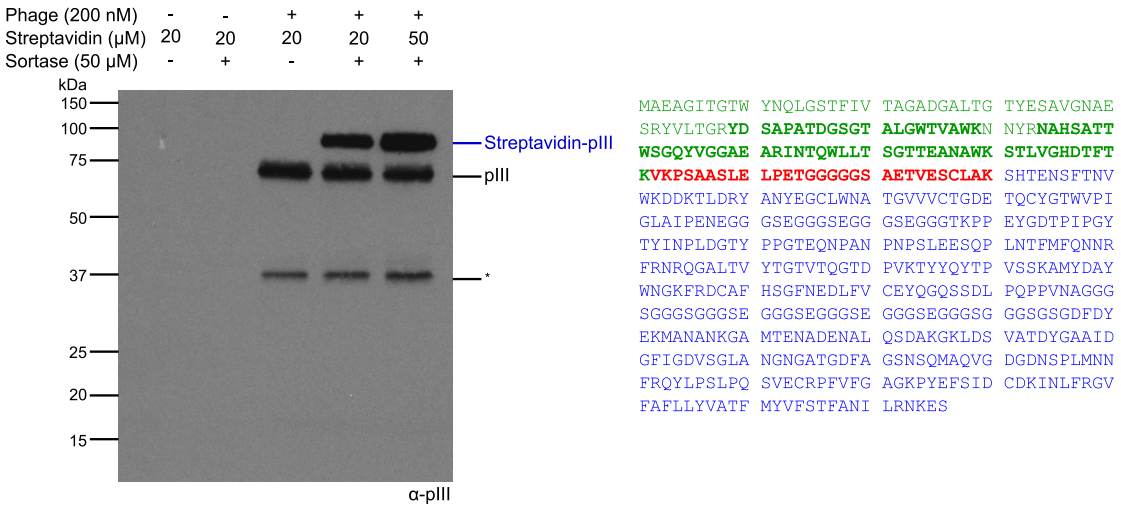


Figure 2.10: **pIII labeling with streptavidin**. G_5-pIII phage (there are five copies of pIII/ phage particle) was incubated with $SrtA_{aureus}$ and streptavidin containing a C-terminal LPETG motif in each monomer. The reactions were monitored by SDS-PAGE under reducing conditions followed by immunoblotting using an anti-pIII antibody. The molecular weight markers are shown on the left. The unidentified anti-pIII reactive protein (*) is attributed to proteolyzed pIII. The identity of the streptavidin-pIII fusion product was determined by mass spectrometry. The amino acid sequences of pIII and streptavidin monomer are shown in blue and green, respectively. The peptides identified are highlighted in bold. The tryptic peptide comprising the streptavidin C-terminus, followed by the $SrtA_{aureus}$ cleavage site, fused to the N-terminal glycines of pIII is shown in red.

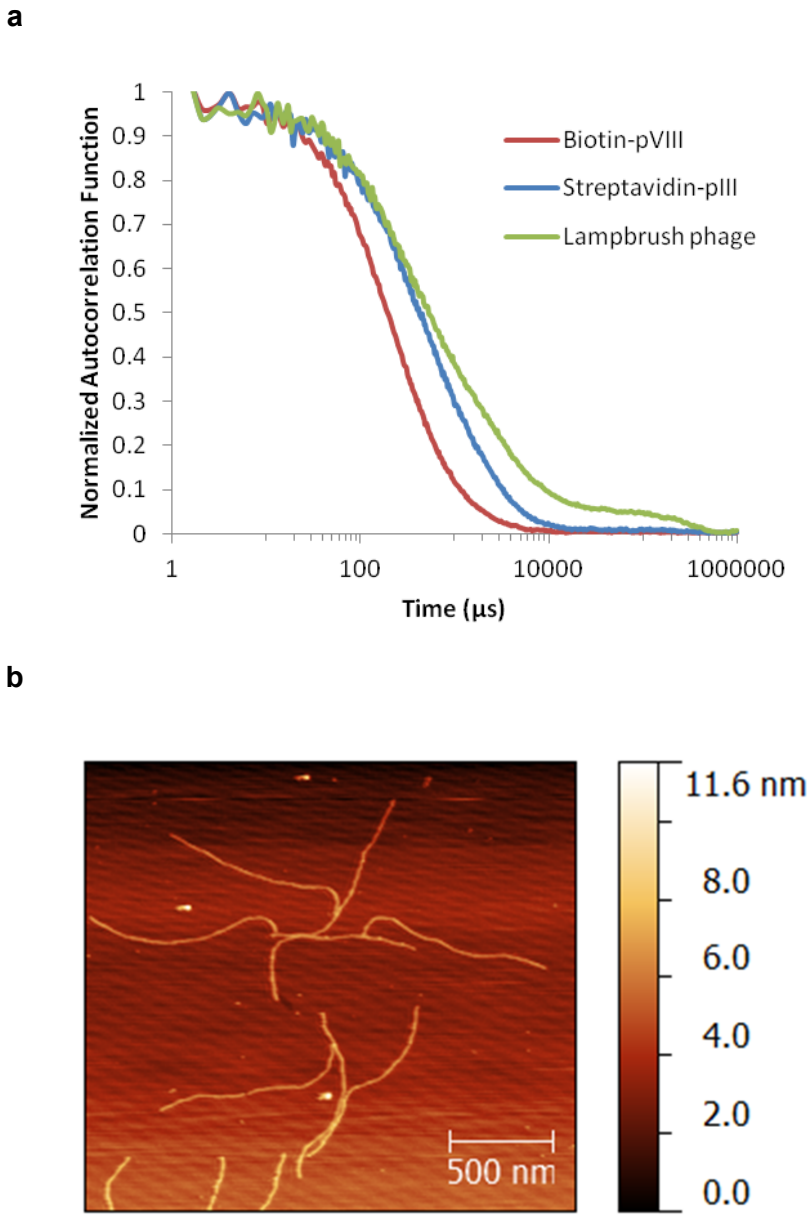


Figure 2.11: **Characterization of lampbrush phage structure**. Upon labeling of the N-terminus of pIII with streptavidin and of the N-terminus of pVIII with biotin using sortase-mediated reactions, the phage were mixed (see Methods section for details). The resulting product was visualized by dynamic light scattering (a) and by atomic force microscopy (b).

concept, a phage variant that contains a double alanine at the N-terminus of pVIII and the pentaglycine motif at the N-terminus of pIII was generated (this construct is referred to as G_5-pIII-A_2-pVIII). Conditions were optimized to label each of these proteins in a site-specific manner. Because such dual-labeled phage could be a useful tool to sort cells by FACS (see below and Discussion section), we here provide the proof-of-concept by labeling the body of phage with a fluorophore and the tip of phage with a cell-targeting moiety.

pVIII was labeled with a K(TAMRA)-LPETAA peptide and purified using PEG/NaCl precipitation to remove free peptide and sortase (Fig. 2.13a). A fluorescent 10kDa protein, corresponding to pVIII, was the only polypeptide detected in the complete reaction. This confirmed successful labeling and site-specificity of $SrtA_{pyogenes}$. The pIII of this fluorescent phage was then incubated with $SrtA_{aureus}$ and a 15kDa single domain antibody, VHH7, modified with a C-terminal LPETG motif. VHH7 recognizes murine Class II MHC products (the development and expression of VHH7 will be described elsewhere). Attachment of VHH7 to pIII was monitored by immunoblot using an anti-pIII antibody (Fig. 2.13b). Comparing the signal intensities of VHH7-pIII 90kDa polypeptide and of pIII, we estimated that on average 2-3 VHH7 molecules are attached per phage particle, a number similar to what can be obtained when screening phagemid libraries of pIII fusions by panning [77, 78].

2.2.6 Flow cytometry experiments using fluorescent phage

Fluorescent phage has been used for targeted staining *in vivo* [29, 79] as well as flow cytometry experiments [80]. However, these have been performed with short peptide phage display libraries. The ability to label phage with a large number of fluorophores that are site-specifically attached to pVIII is a tool useful for selecting phage of interest from phage display libraries of large moieties (such as antibodies)

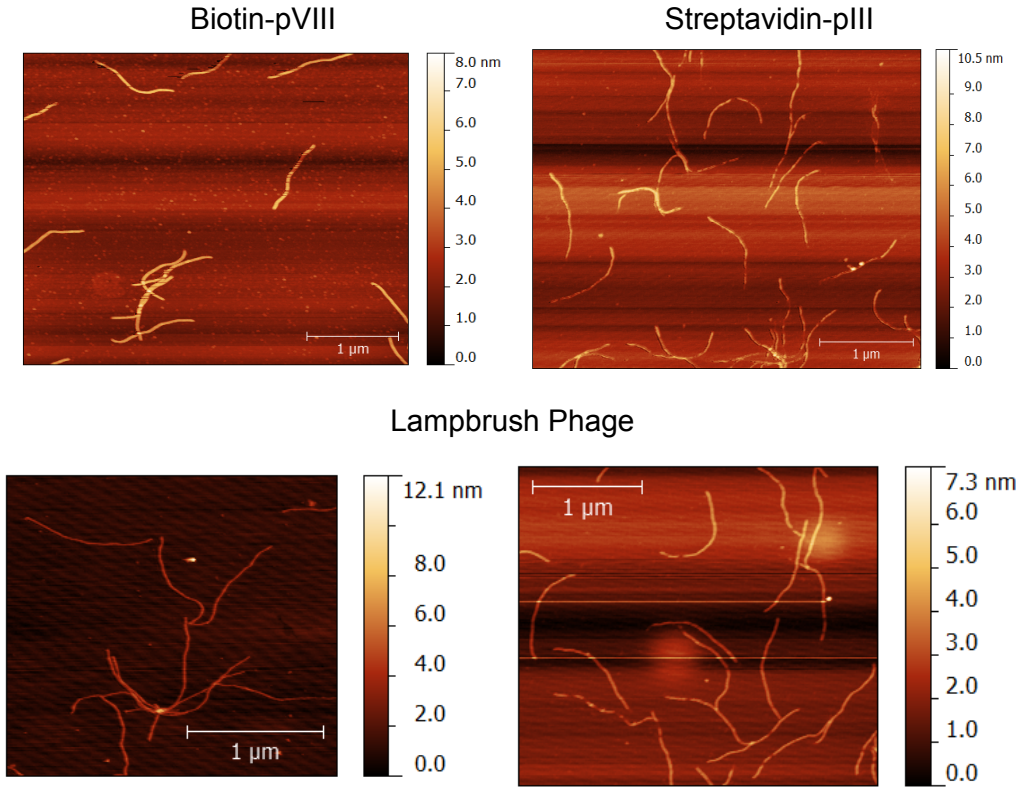


Figure 2.12: **AFM characterization of lampbrush phage structure**. Phage with the N-terminus of pIII labeled with streptavidin and phage with the N-terminus of pVIII conjugated to biotin were created using sortase-mediated reactions. The phage preparations were visualized by AFM before (top right and top left panels) and after mixing (bottom panels; see Methods section for details).

by fluorescence. With libraries of this type, less specific labeling methods can alter the displayed moiety. To provide proof-of-concept that fluorescent phage can be used for this purpose, we tested the ability of the dual labeled phage - containing TAMRA fluorophore sortagged onto pVIII and VHH7 onto pIII - to stain B cells. As a negative control, we used a fluorescent phage containing an anti-GFP VHH attached to pIII [81]. An average yield of 2.5 antibodies per phage virion was achieved for both VHH7 and anti-GFP VHH as determined by densitometric analysis.

Mouse lymphocytes obtained from lymph nodes were stained for B cells using a fluorescent Pacific Blue anti-mouse B220 antibody and incubated with phage-VHH7, phage-anti-GFP, or non-targeted phage. All phage preparations were similarly labeled with TAMRA on pVIII. After removal of unbound materials by washing, cells were subjected to flow cytometry (Fig. 2.14). When stained with phage-VHH7, we detected an increase in cells double positive for TAMRA and the B cell marker compared to non-specific staining with phage-anti-GFP or non-targeted phage. Staining of cells with phage-VHH7 was vastly superior to VHH7 directly conjugated to TAMRA, as only a few double positive cells were detected when incubated with an equivalent amount of the latter (Fig. 2.14).

2.3 Discussion

We show that sortase-mediated reactions overcome many of the limitations of current methods to functionalize M13 capsid proteins. The main body and both ends of the viral capsid can be functionalized with substituents that cannot be encoded genetically (such as biotin and fluorophores), and we can also install properly folded and assembled proteins (such as GFP and streptavidin) in a manner that could easily be extended to oligomeric proteins as well.

One of the major challenges has been the modification of the major capsid

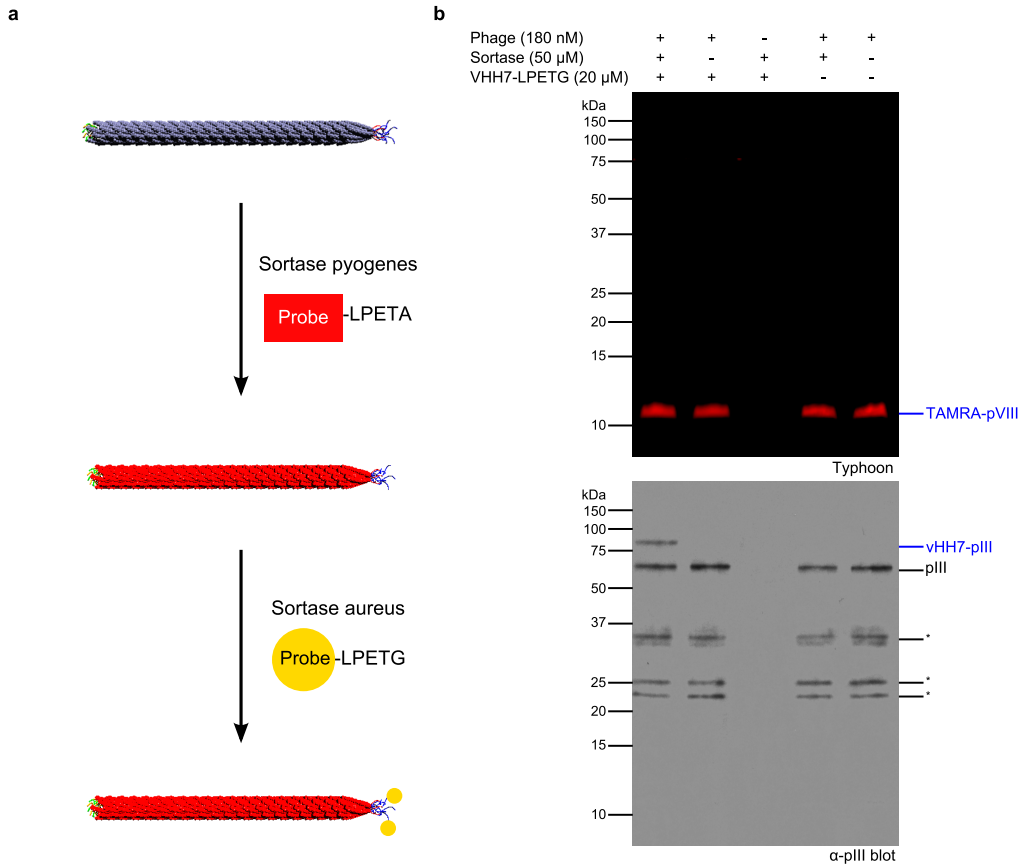


Figure 2.13: **Dual labeling of phage using orthogonal $\mathbf{SrtA_{pyogenes}}$ and $\mathbf{SrtA_{aureus}}$.** Schematic representation of the strategy used to couple two different moieties to two different capsid proteins (a). Labeling of pVIII with a K(TAMRA)-LPETAA peptide mediated by $SrtA_{pyogenes}$ was followed by labeling of pIII with a single domain antibody directed to Class II MHC as a cell targeting moiety and $SrtA_{aureus}$ (see Methods section for details). The final product was analyzed by fluorescent scanning imaging to visualize labeling of pVIII, followed by immunoblotting using an anti-pIII antibody to monitor the efficiency of labeling (b). There are five copies of pIII/ phage particle. The asterisks indicate unidentified anti-pIII reactive proteins, which probably correspond to proteolytic fragments of pIII.

protein pVIII. Using sortase, labeling efficiencies were greater than those obtained genetically (Table 2.1). In the past, biotinylated phage has been produced by display of the biotin acceptor peptide (BAP) [82], a 15-amino acid sequence. Peptides similar in size have been displayed at no more than 400-700 copies per phage, with the efficiency being sequence-dependent [64]. Here we attach 1350 biotin molecules on average per phage particle, a great improvement in the display of a small molecule. Moreover, because the peptide substrate for sortase can be modified with peptides, proteins, fluorophores, etc. [46–50], phage can be decorated with a wide range of substituents. As far as display of proteins is concerned, proteins similar in size to GFP have been incorporated at fewer than one copy per phage on pVIII using a phagemid system [62]. Using sortase, we display ∼90 GFP molecules on average per phage particle.

Table 2.1: **Labeling efficiency for each of the phage coat proteins using sortase**.

Minor Capsid Proteins				
Capsid Protein		**Probe**	**Efficiency**	
pIII		Biotin	68±9%	
pIII		GFP	56±2%	
pIX		Biotin	73±2%	
pIX		GFP	74±1%	
Major Capsid Protein				
Capsid Protein	**Probe**	**Optimal Packing**	**Copy Number/Phage Using Sortase**	**Literature**
pVIII	Biotin	2700	1350±90	400-700
pVIII	GFP	385	91±20	<1

For the pIII and pIX proteins, we show that every phage can be labeled with multiple copies of the desired peptide/protein (Table 2.1). An advantage of using sortase to covalently attach proteins to phage over genetically engineering pIII directly is that it ensures display of the correct quaternary structure of the protein. This can be inferred from our experiments using streptavidin. The mixing of two phage particles, one containing streptavidin on pIII and the other containing biotin on pVIII

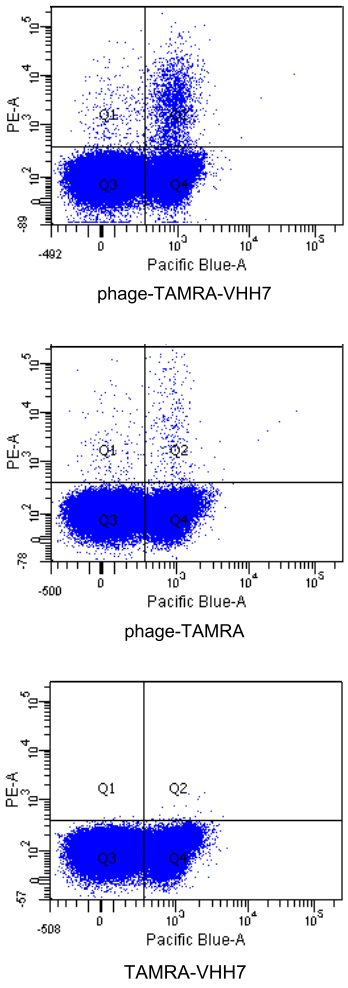


Figure 2.14: **Staining of Class II MHC+ cells using dual labeled phage**. Binding of the dual labeled phage to lymphocytic Class II MHC+ cells was observed by flow cytometry. The Class II MHC+ enriched cell fraction of the lymph nodes of a C57BL/6 mouse was stained for B220 together with the dual labeled phage (phage-TAMRA-VHH7), TAMRA labeled phage (no cell targeting motif, phage-TAMRA), or anti-Class II MHC directly conjugated to TAMRA (TAMRA-VHH7).

results in a novel and complex phage structure. This shows that the streptavidin structure displayed on phage remains fully active and binds biotin.

Sortase enzymes in combination with the streptavidin-biotin pair [83] or in conjunction with click-chemistry can generate novel structures. The ability of patterning and aligning materials on phage or of increasing its surface area is important for the development of new materials. For example, the lampbrush phage structure generated here (Fig. 2.9) may find application in light-sensitive processes where phage branching off the stem could be functionalized to act as antennae to capture light [84].

In addition to N-terminal labeling of single capsid proteins, two capsid proteins were labeled site-specifically on a single phage particle using two orthogonal sortases. This could be explored for panning of antibody libraries displayed on pIII. Due to the exquisite site-specificity of sortase, fluorescent peptides can be added to pVIII without modification of the moiety displayed at pIII. Fluorescent labeling by other chemistries does not easily afford such specificity, especially when displaying a large moiety, such as an antibody fragment. The sensitivity of detection should increase when a phage particle contains many fluorophore groups on pVIII. This is indeed what we observe in our flow cytometry experiments, showing that this strategy greatly enhances the sensitivity of detection. Increased sensitivity would be instrumental in the context of a future panning strategy for detection of rare binding events, whether due to low concentration of the target or low phage concentration.

Modification of pIII and pIX by sortase will be useful first and foremost for material applications, where the physical properties of phage and not its utility as a library vector are of prime concern. Fluorescent modification of pVIII is compatible with the construction and screening of libraries created using pIII genetic fusions. In this case, the site-specificity and yield of the sortase reaction allow the generation of libraries that can be screened directly by fluorescence. Thus, the versatility of the sortase-based labeling strategy described here will enable development of a wide array

of tools, expanding the use of phage either for the creation of new materials or for new biological applications.

2.4 Methods

Generation of M13 phage constructs

The oligonucleotides used to design the different phage constructs are compiled in Supplementary Table A.1. The G_5-pIII phage was engineered by inserting the G5pIIIC and G5pIIINC annealed oligonucleotides into the M13KE vector (New England Biolabs), previously digested with *Eag*I and *Acc*65I restriction enzymes. To construct the A_2G_4-pVIII phage, the M13SK vector [23] was digested with *Pst*I and *Bam*HI restriction enzymes and the A2G4pVIIIC and A2G4pVIIINC annealed oligonucleotides were inserted. To engineer the G_5HA-pIX construct, the 983 vector was used (Ghosh, D. *et al.*, unpublished). This vector was created by refactoring the M13SK vector so the pIX and pVII genes are not overlapping. Upon digestion of this vector with *Sfi*I, the annealed G5HApIXC and G5HApIXNC oligonucleotides were inserted. The G_5-pIII-A_2-pVIII phage construct was created using a modified M13SK vector [23], which has a DSPHTELP sequence on pVIII and a biotin acceptor peptide (GLQDIFEAQKIEWHE) on pIII. Five N-terminal glycines were added to pIII following the above strategy described for G_5-pIII phage. The resultant vector was then modified at the N-terminus of pVIII using the QuikChange II site-directed mutagenesis kit (Stratagene) and the pVIIIAADSPH oligonucleotide pair. All the generated phage vectors were transformed into the XL-1 Blue bacterial strain, plated in agar top on LB agar plates containing 1mM IPTG, 40μg/mL X-Gal, and 30μg/mL tetracycline. Plaques were selected and DNA was isolated and sequenced to check for the insertion.

For phage amplification, the *E. coli* strain ER2738 (New England Biolabs)

in LB media supplemented with 30μg/mL tetracycline, was infected with phage for at least 12hrs at 37°C. The cultures were centrifuged at 12000g for 20min and the phage was precipitated from the supernatant at 4°C with the addition of 1/5 of the supernatant volume of 20% PEG8000/2.5M NaCl solution. Upon centrifugation at 13500g for 20min, the pellet was resuspended in 25mM Tris, 150mM NaCl, pH 7.0-7.4 (TBS). For further purification, this resuspension was subjected to two rounds of centrifugation/precipitation. The final phage concentration averaged between 10^{13}-10^{14} plaque forming units (pfu) per mL as determined by UV-vis spectrometry [76].

Sortase-mediated reactions

$SrtA_{pyogenes}$ and $SrtA_{aureus}$ were expressed and purified as described [47, 85]. Sortase reactions were performed as indicated in the figures. A typical sortase reaction with $SrtA_{aureus}$ included 200nM phage, 50μM $SrtA_{aureus}$, and 50μM substrate for small peptides or 20μM for proteins. The reactions were incubated for 3hrs at 37°C (for small peptides) or at room temperature (for proteins) in TBS with 10mM $CaCl_2$. $SrtA_{pyogenes}$-mediated reactions included 8nM phage, 50μM $SrtA_{pyogenes}$, and 20μM substrate, incubated for 3hrs at 37°C in TBS. Where indicated, phage was purified by PEG 8000/NaCl precipitation after diluting the reactions with TBS such that the substrate concentration was below 600nM.

For the flow cytometry experiments, the G_5-pIII-A_2-pVIII phage construct was labeled with K(TAMRA)-LPETAA on pVIII. The resultant labeled phage was purified by PEG8000/NaCl precipitation, resuspended in TBS, and split into three parts. One part remained unlabeled, and the other two were labeled with either VHH7.LPETG or anti-GFP.LPETG on pIII. As assessed by the anti-pIII antibody, a yield of 2.5 antibody molecules per virion was achieved in both cases.

The yield of the sortase-mediated biotinylation reactions was determined using biotinylated GFP as a standard. This was prepared labeling GFP - comprising a

LPETG at its C-terminus - with a biotin group using $SrtA_{aureus}$ (GFP.LPETGGGK (biotin)) [85]. Known amounts of the purified GFP.LPETGGGK(biotin) standard and varying volumes of the phage labeling reactions were loaded onto the same SDS-PAGE gel and analyzed by immunoblot using streptavidin-HRP (GE Healthcare). The signal obtained in the phage labeling reactions was compared with the signal derived from the GFP.LPETGGGK(biotin) calibration curve allowing us to infer the amount of phage protein labeled in the reaction. To calculate the labeling efficiency, the amount of labeled protein was divided by the amount of total phage protein loaded into the gel. The phage concentration was determined by UV-vis spectrometry and it was assumed that there were 2700 copies of pVIII, 5 copies of pIII, and 5 copies of pIX per phage particle.

To determine the yield of GFP-pVIII phage labeling, unincorporated GFP and sortase was removed from phage by PEG8000/NaCl precipitation. Varying volumes of GFP-pVIII phage and known amounts of GFP were loaded onto the same SDS-PAGE gel and analyzed by immunoblot using an anti-GFP-HRP antibody (Santa Cruz Biotechnology). The signal of the GFP-pVIII fusion protein was compared to the signal of the GFP calibration curve as described for the biotinylation reactions. For GFP-pIII and GFP-pIX labeling, the signal of the fusion protein was compared to the input amount of pIII or pIX as detected by anti-pIII (New England Biolabs) or anti-HA (Roche) antibodies, respectively. For GFP-pIII, the input signal consisted of only intact pIII molecules and lower molecular weight anti-pIII reactive proteins were not included. These proteins can be attributed to proteolyzed pIII [86]. Because the anti-pIII antibody recognizes the C-terminus of the protein, these fragments cannot be labeled using $SrtA_{aureus}$. In all cases the blots were scanned and densitometric analysis was performed using the ImageJ program (National Institutes of Health). The labeling yield was averaged over three independent reactions with three aliquots from each reaction analyzed. The standard deviation of the reactions was calculated

from the averages of the three independent reactions.

Dynamic light scattering (DLS)

DLS measurements were obtained with a Beckman Delsa-Nano C Particle Analyzer (Beckman Coulter Inc). Phage mixtures were diluted to $\sim 10^{11}$pfu/mL in 1mL of water and loaded into a cuvette. Samples from each experiment were measured in triplicate and the results were averaged by cumulant analysis. Autocorrelation functions were used as a direct comparison of aggregation because aggregates have a slower Brownian motion causing the signal correlation to be delayed to longer relaxation times.

Atomic force microscopy (AFM)

Phage preparations were diluted to a concentration of $\sim 10^{11}$pfu/mL, and 100μL of this mixture were deposited on a freshly cleaved mica disc. AFM images were taken on a Nanoscope IV (Digital Instruments) in air using tapping mode. The tips had spring constants of 20-100N/m driven near their resonant frequency of 200-400kHz (MikroMasch). Scan rates were approximately 1Hz. Images were leveled using a first-order plane fit to remove sample tilt.

Flow cytometry analysis

C57BL/6 mice were purchased from Jackson Labs. Animals were housed at the Whitehead Institute for Biomedical Research and were maintained according to guidelines approved by the Massachusetts Institute of Technology (MIT) Committee on Animal Care.

Lymph nodes were isolated from 6-8 week old C57BL/6 mice and crushed through a 40μM cell strainer. Cells were washed once with PBS, resuspended at

$2x10^7$ cells per mL, aliquoted at $\sim 1x10^6$ cells per sample, and incubated with staining agents in 5% milk in PBS for 1hr at room temperature. 10^{11} VHH7 molecules and 10^{11} anti-GFP molecules either directly conjugated to TAMRA using $SrtA_{aureus}$, or covalently attached to phage ($5x10^{10}$ phage particles of VHH7-G_5-pIII-TAMRA-A_2-pVIII or anti-GFP-G_5-pIII-TAMRA-A_2-pVIII, see *Sortase-mediated reactions* section) were incubated with the cells. The same amount of non-targeted fluorescent phage particles (i.e., G_5-pIII-TAMRA-A_2-pVIII) was used as a negative control. B cells were stained with Pacific Blue anti-mouse B220 (BD Pharmingen, clone RA3-6B2). Upon staining, the cells were centrifuged at $170g$ for 5min, washed with PBS three times, and resuspended in 500μL of PBS. Flow cytometry was performed using a FACSAria (BD). 100,000 events were collected for each sample.

Miscellaneous

Expression and purification of GFP.LPETG.His_6 and GFP.LPETA.His_6, were performed as described [47]. Identification, characterization, expression, and purification of VHH7.LPETG.His_6 will be described elsewhere. Streptavidin was cloned as a streptavidin.LPETG.HAtag.His_6 fusion protein using the template Addgene 20860 [87], and expressed as a soluble tetrameric streptavidin [87]. Purification was performed following the same protocol used for GFP [47]. Sortase reactions were analyzed on 4-12% Bis-Tris SDS-PAGE gels with MES running buffer except for Fig. 2.10 which was analyzed on a 12% Laemmli SDS-PAGE gel.

The K(biotin)-LPETGG, K(biotin)-LPETAA, K(TAMRA)-LPETAA, and GG-GK(biotin) peptides were obtained from the Swanson Biotechnology Center. For mass spectrometry, the protein bands of interest were excised, subjected to protease digestion, and analyzed by electrospray ionization tandem mass spectrometry (MS/MS). Fluorescent gel images were obtained using a variable mode imager (Typhoon 9200; GE Healthcare).

Chapter 3

C-terminal sortase-mediated labeling of M13 bacteriophage

3.1 Introduction

M13 bacteriophage is composed of five capsid proteins (pIII, pVI, pVII, pVIII, and pIX). In the process of phage display, genetic fusions of these capsid proteins with peptide/protein libraries have allowed for the selection of biomolecules that bind organic [25, 26] and inorganic materials [27]. Engineered phage displaying these selected biomolecules can form the framework for devices with applications in the biological and material sciences [35, 38, 39].

Phage display can be achieved through two methods. The first uses phagemids which are plasmids that encode a mutant form of a capsid protein. Helper phage infects bacteria containing the plasmid and the replicated phage contains a mixture of the mutant form as well as the capsid protein of the helper phage creating a hybrid virion. The second method is phage vector display where the phage genome is modified directly ensuring that every copy of the capsid protein is modified, but the method has a more strict size limitation to the fusions than phagemids.

Phage display has primarily been performed on the N-termini of the capsid proteins [71], as the C-termini of the phage capsid proteins are buried with the exception of pVI [75]. Although C-terminal display has been achieved with phagemids [69, 70, 88, 89], the fusions have been expressed with low efficiency. For phage vector display, there have been no documented C-terminal fusion proteins. Nonetheless,

C-terminal display is required for certain protein motifs to maintain their function [90, 91]. The installment of a conjugation chemistry motif would allow for the display of these moieties on the surface of phage.

Sortase A enzymes allow modification of the C-terminus of proteins through a chemo-enzymatic reaction with a wide range of functional groups including biotin, fluorophores, and other proteins [46, 48, 50]. Sortase A from *Staphylococcus aureus* ($SrtA_{aureus}$) recognizes an LPXTG substrate sequence [44]. The enzyme cleaves between the threonine and glycine forming a covalent acyl-enzyme intermediate, which is resolved by the attack of an N-terminally exposed oligoglycine sequence.

Sortase-mediated reactions are well-suited to C-terminal labeling of capsid proteins. First, the motif is small which makes it a candidate to be expressed in a phage vector on the C-terminus of the proteins. Second, the labeling method is versatile and capable of attaching both small and large moieties to the end of phage. Third, the reaction is site specific and none of the capsid proteins naturally express the nucelophile or substrate sequences required for a sortase reaction.

Initial attempts at C-terminal display of the sortase substrate motif were unsuccessful. However, display of a C-terminus could still be achieved through a loop architecture that can expose a C-terminus upon proteolysis and remain attached by a disulfide bridge between two cysteine residues. The structure of the loop is important and many of the possible sequences may not form a loop even with two cysteine residues present. Therefore, we used inspiration from cholera toxin (Ctx), which contains an accessible loop in its structure. The Ctx loop was engineered in order to label the A1 subunit of cholera toxin with sortase by displaying the LPXTG sequence [48].

We transplanted the sortaggable loop from the cholera toxin subunit to pIII creating an architecture for C-terminal phage display and C-terminal sortase labeling on phage. We demonstrate that the Ctx loop architecture can conjugate small and

large moieties using sortase expanding the moieties that can be displayed on pIII. With this ability, we created end-to-end dimer structures linking our loop architecture with phage containing the sortase nucleophile motif.

3.2 Results

3.2.1 C-terminal phage vector display of the sortase substrate motif

Table 3.1: **Restriction enzymes used for C-terminal phage vector display**

Capsid Protein	**Enzymes**
pIII	*Bsp*EI & *Bcl*I
pVI	*Aat*II & *Age*I
pIX	*Spe*I & *Bsp*HI

To test if C-terminal phage vector display was possible, we genetically fused the sortase substrate motif to the C-terminus of the pIII, pVI, and pIX proteins. Near the C-termini end of the genes encoding these capsid proteins, we introduced silent mutations to insert restriction sites that cut a single time in the phage genome. The restriction sites for each capsid protein are listed in Table 3.1. The pVII and pVIII were not investigated as the C-termini of these proteins are less exposed than the other capsid proteins [75]. Using the inserted restriction sites, we add an LPETGG motif followed by an HA tag to the C-termini of the capsid proteins. The HA tag was added in order to determine whether the C-terminus was exposed to participate in the sortase reaction. By inserting no linker, GGGS, and $(GGGS)_3$ upstream, we extended and altered the flexibility of the sortase motif. For each of these insertions, we analyzed the ligation reaction by PCR to check that a phage vector containing the new insert was present (Fig. 3.1). In these PCR reactions, we detected a ~500bp

DNA species when the primers that matched the insertion were included. Despite the confirmed presence of the inserted DNA sequence, we found no phage that contained the desired modification.

3.2.2 Display of a sortaggable loop on pIII

We circumvented the inability to perform C-terminal display by introducing a sortaggable loop onto pIII. Previously, a sortaggable loop had been created on cholera toxin A1 subunit [48]. This loop contained an LPXTG recognition motif for $SrtA_{aureus}$ followed by an HA tag and a trypsin cut site. The trypsin cut site was used to open the loop to improve labeling efficiency. The HA tag was added to monitor sortase activity as it is released by sortase in an opened loop. We engineered the phage genome to display the loop fused to the N-terminus of pIII. Sortagging with a GGGK(TAMRA) probe was observed only when trypsin was added to the reaction (data not shown). However, trypsin digested pIII in addition to opening the loop. Despite all of our attempts, we did not find a concentration of trypsin compatible with opening the loop efficiently while simultaneously maintaining the structure of pIII.

3.2.3 C-terminal labeling of Factor Xa loop

In order to avoid proteolysis other than opening the loop, the trypsin cut site was modified by mutagenesis to a Factor Xa site as this protease does not digest pIII. Using this LoopXa-pIII construct, the loop can be cleaved completely without disrupting the pIII structure. A schematic of the labeling method is shown in Fig. 3.2.

We labeled the loop with a GGGK(TAMRA) peptide using $SrtA_{aureus}$ (Fig. 3.3). We analyzed the reactions by SDS-PAGE under both reducing and non-reducing conditions followed by fluorescent imaging and Coomassie staining. We observed a

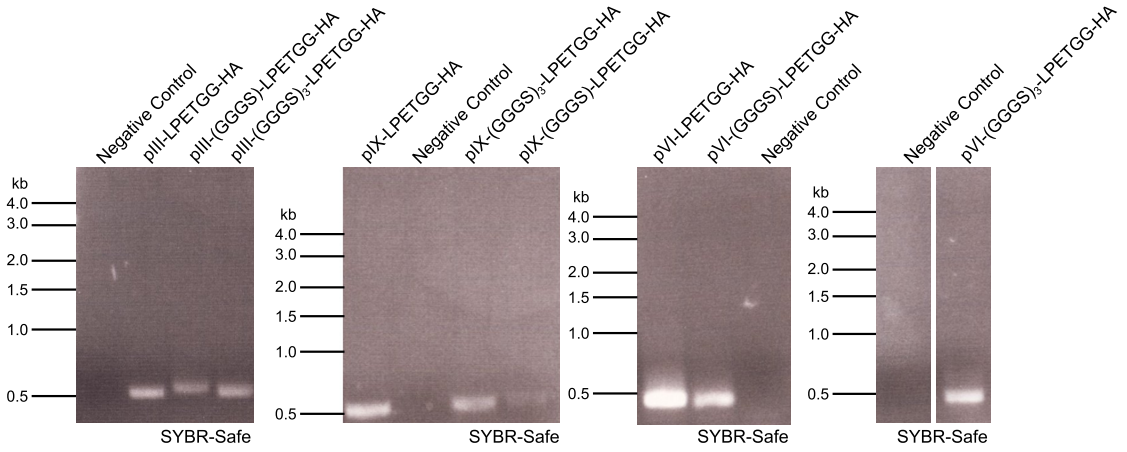


Figure 3.1: **C-terminal display on pIII, pVI, and pIX**. DNA insertions encoding LPETGG-(HA), GGGS-LPETGG-(HA), and $(GGGS)_3$-LPETGG-(HA) were inserted genetically at the C-terminus of pIII, pIX, and pVI. The ligation reactions of these inserts were analyzed by PCR using one of the insertion oligonucleotides from the ligation and a second primer matching an unmodified part of the phage vector oligo (see Methods for details) to determine whether the inserts had been incorporated into the genome.

60kDa fluorescent protein when LoopXa-pIII phage, GGGK(TAMRA), $SrtA_{aureus}$, and Factor Xa were mixed together in non-reducing conditions. The same 60kDa polypeptide corresponds to the LoopXa-pIII protein, which was observed by Coomassie staining (Fig. 3.3).

To demonstrate labeling with a large moiety, we labeled the loop with cholera toxin subunit B (CtxB). The B subunit forms a 58kDa pentamer [92] and each subunit was expressed with a G_3 nucleophilic motif for $SrtA_{aureus}$. In the presence of SDS and heat, the pentamer can be converted to its monomer form. LoopXa-pIII phage was incubated with Factor Xa, $SrtA_{aureus}$, and G_3-CtxB for 5hrs at room temperature. We monitored the reactions by SDS-PAGE under reducing and non-reducing conditions followed by immunoblot with an anti-pIII and anti-CtxB antibody (Fig. 3.4). When phage, $SrtA_{aureus}$, Factor Xa, and G_3-CtxB were mixed in non-reducing conditions, we detected a 75kDa anti-pIII and anti-CtxB reactive protein consistent with the size of the pIII-CtxB fusion. This identity of this protein was confirmed by mass spectrometry (Fig. 3.4 and Supplementary Fig. A.4)

3.2.4 Building of end-to-end phage structures

Phage has been used to biotemplate materials [28, 35, 36], but the ability to control and link the virions can expand the functionality of the phage to form patterned multi-material devices. One such pattern is the end-to-end structure where two phage can be linked pIII to pIII. Using the loop architecture and $SrtA_{aureus}$, we formed phage to phage linkages as shown in Fig. 3.5.

In order to form these structures, we used G_5-pIII phage containing the nucleophile motif for $SrtA_{aureus}$ [52]. We incubated the LoopXa-pIII phage with $SrtA_{aureus}$, Factor Xa, and G_5-pIII phage for 24hrs with 10mM $CaCl_2$. Due to the slow diffusion of the nucelophile and substrate in this reaction, we did not observe any linkages. In order to improve the chances of linkages, we extended the incubation time to

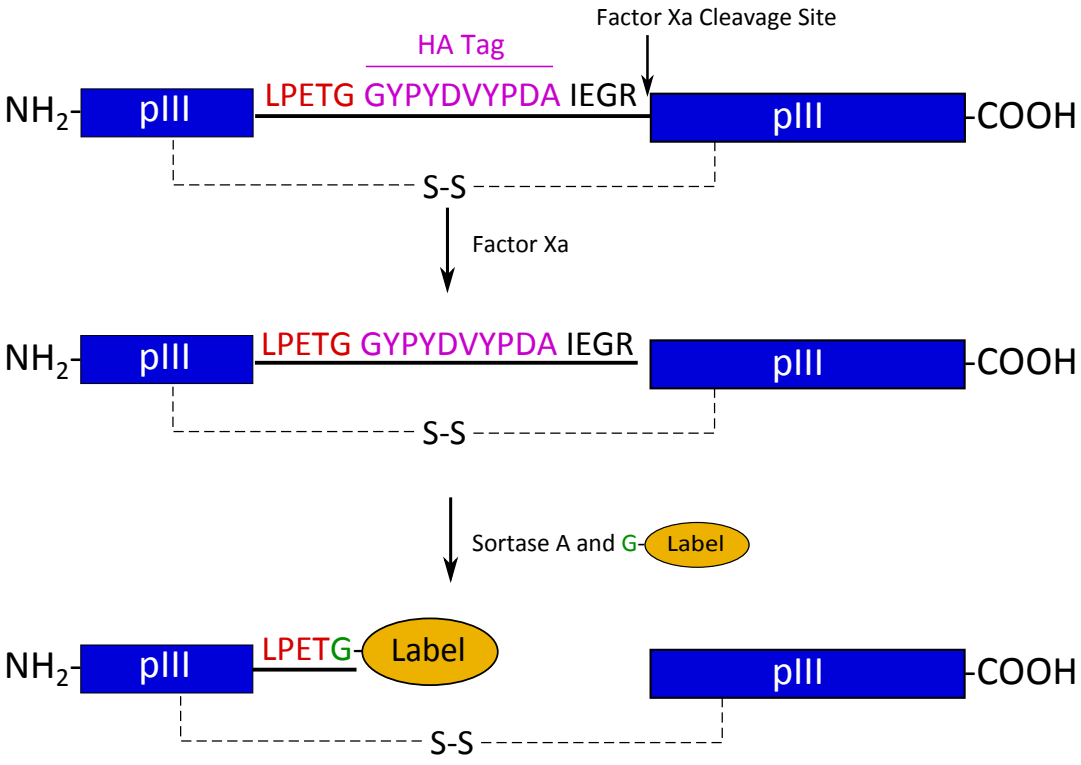


Figure 3.2: **Schematic of sortagging LoopXa-pIII phage**. A loop structure from cholera toxin subunit A was fused to pIII. The loop contains the sortase substrate motif (red), an HA tag (purple), and Factor Xa cleavage site (black). The structure is held together by a disulfide bridge (dotted line).

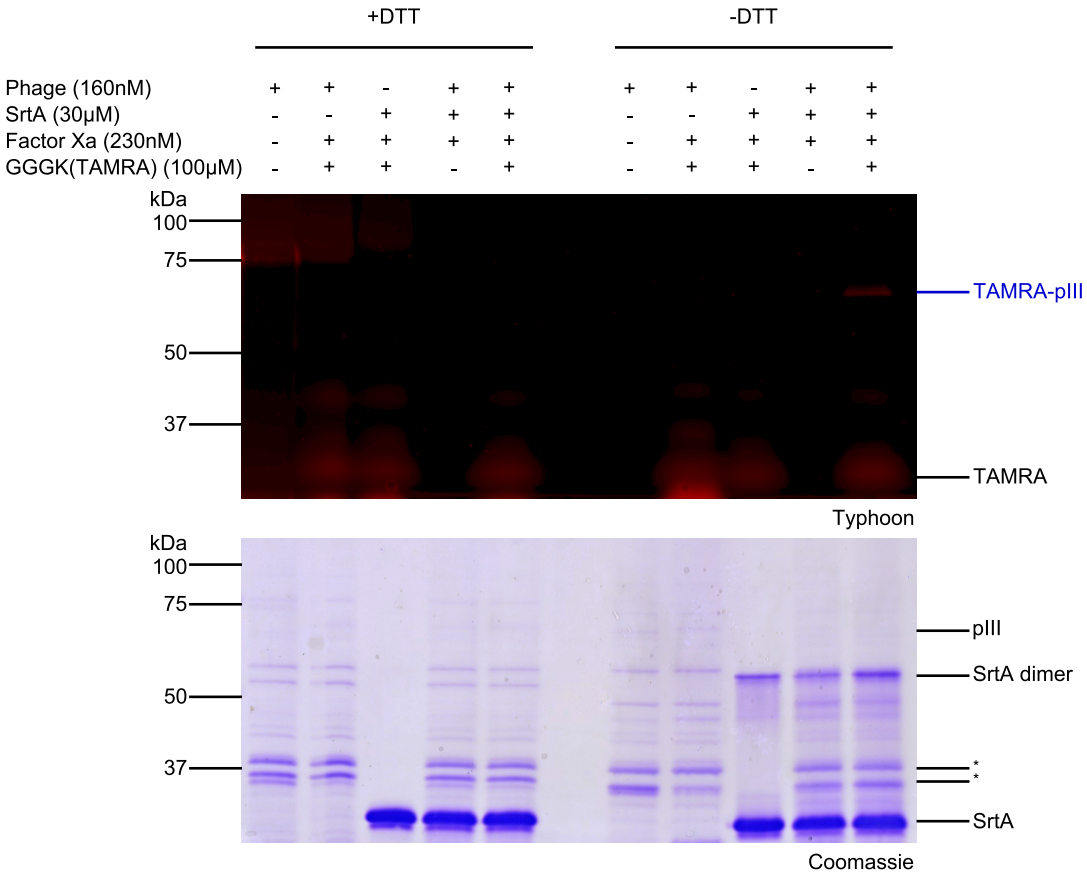


Figure 3.3: **Labeling of pIII loop with TAMRA**. LoopXa-pIII phage (there are five copies of pIII /phage particle) was incubated with Factor Xa, GGGK(TAMRA), and $SrtA_{aureus}$ at room temperature for 5hrs. The reactions were resolved in SDS-PAGE under reducing and non-reducing conditions followed by fluorescent imaging (top) and staining with Coomassie (bottom). The molecular weight markers are shown on the left. The unidentified proteins (*) are most likely *E. coli* proteins from the phage preparation.

60hrs. After incubation, we removed sortase using PEG8000/NaCl purification and resuspended the phage in water. We visualized the purified phage by atomic force microscopy (AFM) (Fig. 3.6a). Linkages between the head of two phages were observed in structures that measured nearly 2μm in length. We estimated that 3% of phage visualized were in the dimer form. To determine that the linkages were the result of sortase and not disulfide bonds formed between cysteines on two different phage, we analyzed G_5-pIII phage, LoopXa-pIII phage, G_5-pIII with LoopXa-pIII phage, LoopXa-pIII phage and $SrtA_{aureus}$ by AFM (Fig. 3.7). We did not observe end-to-end interactions between phage in any of these samples.

We investigated these dimer phage structures by dynamic light scattering (DLS) to monitor the change in particle size (Fig. 3.6b). The total population did not show a change in hydrodynamic radius, but in the normalized autocorrelation function (ACF), we saw a polydisperse population with a second population exhibiting longer relaxation times consistent with structures larger than single phage. We analyzed G_5-pIII and LoopXa-pIII phage by DLS, and we found their ACFs exhibited exponential decay consistent with a monodisperse population of single phage (Fig. 3.6b). To determine that the dimers were sensitive to reducing agents, we added the reducing agent dithiothreitol (DTT) to the pIII-pIII dimer sample. By DLS, we did not detect the population with longer relaxation time and the ACF showed exponential decay similar to those observed for single phage (Fig. 3.6b).

3.3 Discussion

The Ctx loop provides the architecture for C-terminal phage vector display of short peptides such as the sortase substrate motif. We demonstrate that the display of this motif makes it possible to attach diverse moieties at the C-terminus site-specifically. These conjugations demonstrate the versatility of installing the sortase

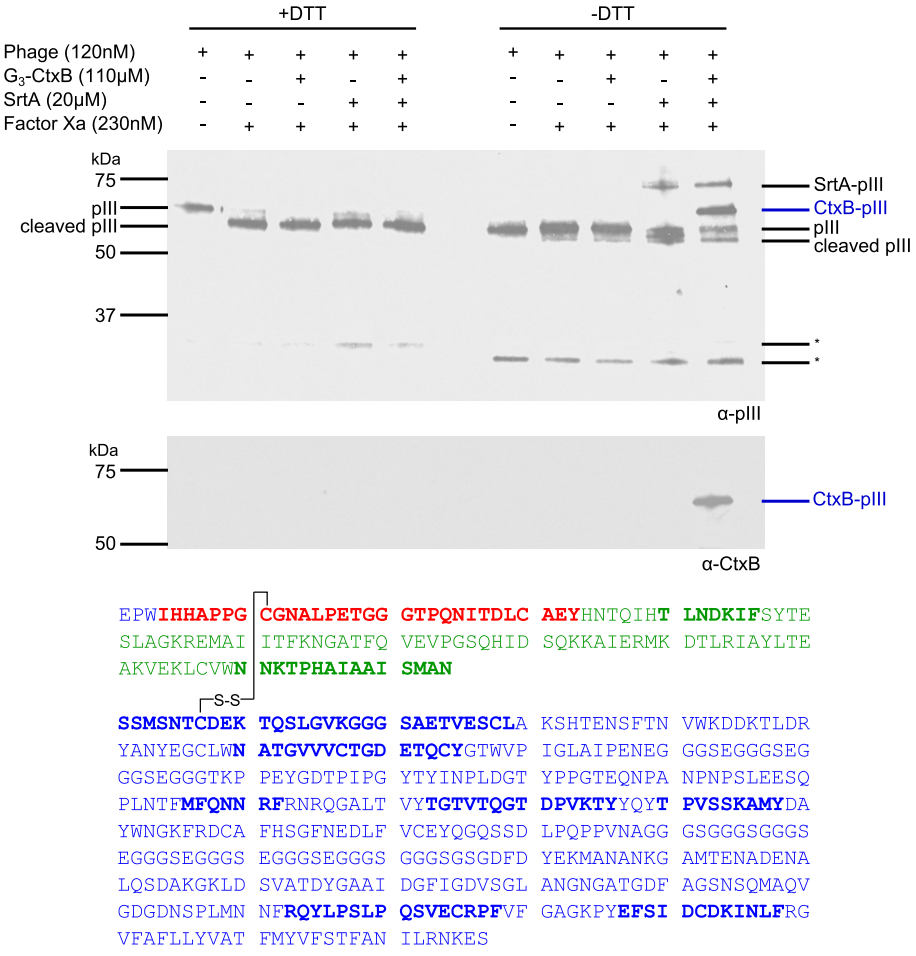


Figure 3.4: **Labeling of pIII loop with CtxB**. LoopXa-pIII phage (there are five copes of pIII /phage particle) was incubated with Factor Xa, G_3-CtxB, and $SrtA_{aureus}$ at room temperature for 5hrs. The reactions were resolved in SDS-PAGE under reducing and non-reducing conditions followed by immunoblot with an anti-pIII antibody and anti-CtxB antibody. The molecular weight is indicated on the left. The unidentified anti-pIII reactive protein (*) is most probably a proteolytic fragment of pIII. The identity of the pIII-CtxB fusion product was determined by mass spectrometry. The amino acid sequences of the LoopXa-pIII and CtxB are shown in blue and green respectively. The disulfide bridge in the loop is shown in black. The peptides identified are highlighted in bold. The chymotryptic peptide comprising the LoopXa-pIII C-terminus followed by the $SrtA_{aureus}$ cleavage site fused to the N-terminal glycines of CtxB is shown in red.

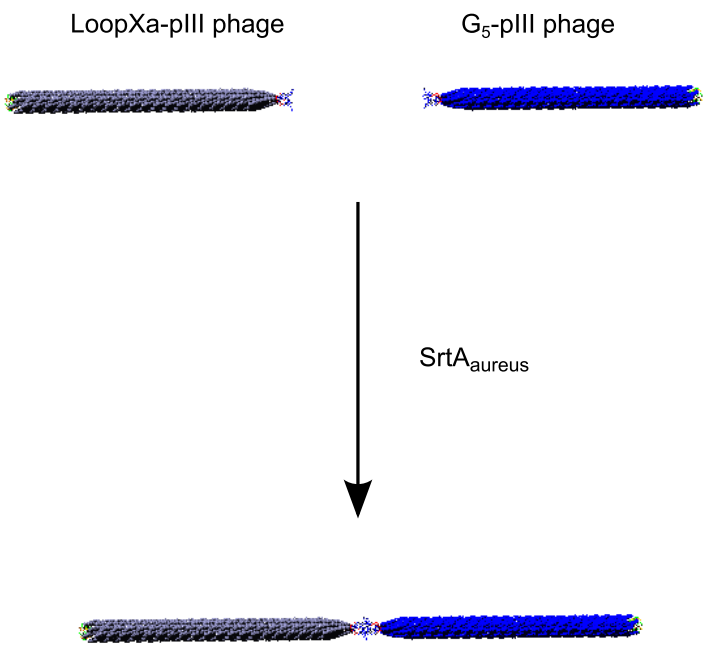


Figure 3.5: **Schematic for end-to-end phage dimers**. Schematic representation of the strategy used to build end-to-end phage structures.

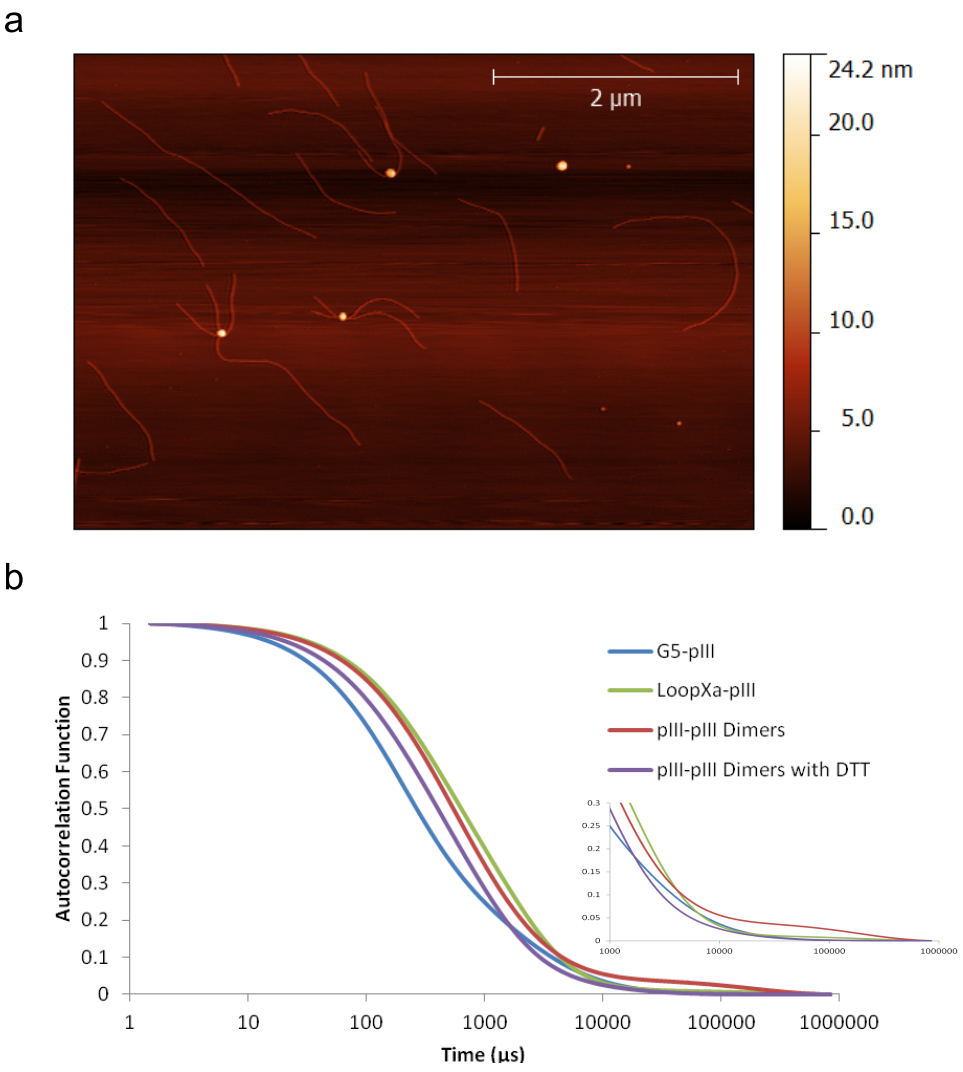


Figure 3.6: **Characterization of end-to-end phage dimers**. G_5-pIII phage, LoopXa-pIII phage, Factor Xa, and $SrtA_{aureus}$ were incubated at room temperature for 60hrs and purified by PEG8000/NaCl (see Methods section for details). The resulting product was visualized by atomic force microscopy (a) and dynamic light scattering with and without reducing agents (b).

motif as we attach two moieties that cannot be genetically encoded: the fluorophore TAMRA and pentameric protein CtxB. For proteins, our sortase labeling method confers the advantage of properly folding the protein separate from the environment of the phage assembly, which is unavoidable in both phagemid and phage vector display.

Phage has been used as a nucelophile in sortase reactions [52], but the loop architecture extends the range of objects that can be attached using sortase. Proteins requiring their C-termini to be exposed for activity can now be displayed. Targeting motifs such as those for the endoplasmic reticulum and peroxisomes, which require the exposure of their C-terminus [90, 93], can remain exposed when displayed on the phage.

The loop architecture is a framework for screening C-terminal peptide libraries. We believe these libraries can be useful in the investigation of PDZ domain interactions, which are more prevalent at the C-terminus of proteins [94]. Currently, T7 phage display is commonly used to probe these interactions [95], but with 415 copies of the modified protein per particle, high avidity interactions are identified. A C-terminal peptide library on pIII would have only 5 copies and higher affinity peptide targets could be identified. Although cysteine constrained libraries are available, the loop architecture can accommodate larger insertions and the protease site allows for the display of a C-terminus.

The architecture also provides a release mechanism through the disulfide bond. This feature could be advantageous for the delivery of cargo to cells. Phage has been used for delivery and can be internalized into mammalian cells [96]. Once inside the cell, the reducing environment breaks the disulfide bridge in the loop and release cargo that was attached using the sortase motif. This release event may be critical to avoiding trafficking and the digestion of cargo in the lysosomes where the phage has been observed to traffic [97]. Beyond delivery of cargo, the disulfide bond provides

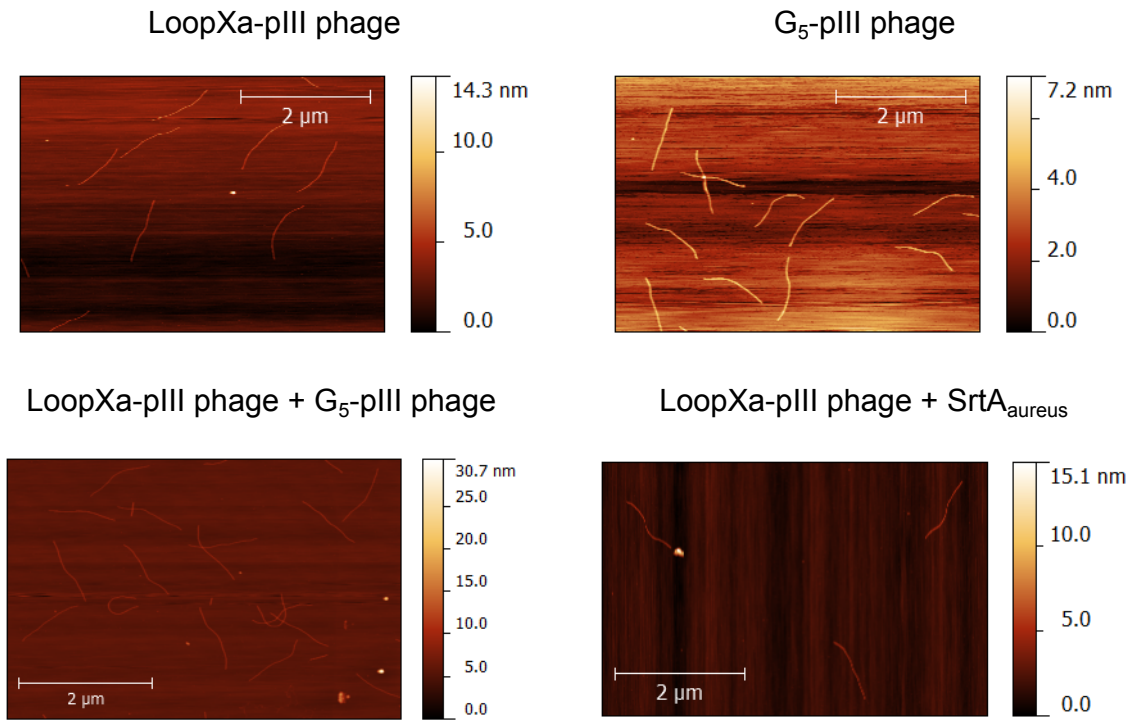


Figure 3.7: **Characterization of end-to-end phage dimers by AFM**. LoopXa-pIII phage and G_5-pIII phage were incubated with $SrtA_{aureus}$ to form phage dimers. The LoopXa-pIII phage and G_5-pIII phage were visualized individually (top panels). The two mixed and the LoopXa-pIII phage with $SrtA_{aureus}$ were analyzed by AFM (bottom panels).

a system for purification of a ligand from a mixture. A phage displaying a target of interest, whether by phage display or sortase labeling, can interact with a ligand in a mixture which can be extracted using PEG8000/NaCl precipitation. Using a reducing agent, it is possible to separate the ligand from the phage.

Sortase has been used to create novel multi-phage structures [52], and we demonstrate that the loop architecture with sortase-mediated reactions can be used to form phage dimers. These connections demonstrate that the sortase-mediated reactions can connect two megadalton sized moieties. Previously, these end-to-end linkages have been made using ordered phage at high concentrations, biotin-streptavidin, or leucine zippers [7, 98, 99]. Compared to the first, our method can produce dimers at phage concentrations an order of magnitude lower and provide more permanent linkages. For the second, the leucine zippers are expressed using the phagemid which can be low in efficiency compared with phage vector display used in our method.

For biotin-streptavidin linkages, the multivalency of streptavidin can lead to node structures forming branched networks that make it difficult to pattern materials in a linear fashion. Because multiple pIII proteins are present on a single phage, multiple linkages are possible using sortase-mediate reactions, but none were observed when forming our linkages. We attribute this 1-to-1 stoichiometry to the steric hindrance provided by one phage once it has been attached. Our method also removes the addition of a tetrameric streptavidin protein, which creates space between the linked phages and consequently the space between the nucleated materials. Given the space between them acts as an insulator, the reduction of this space is critical for conductance properties in the building of biotemplated nanowires.

While C-terminal display has been achieved previously with phagemid display, the ability to expose a C-terminus in phage vector display allows for a more homogenous display across a phage population, which makes it more ideal for peptide library display. With this framework, we demonstrate the ability to functionalize the loop

through sortase-mediated reactions and link an entire phage to form dimer structures. We envision the architecture described in this work will increase the range of objects that can be displayed on M13 as well as improve the ability to create C-terminal display libraries.

3.4 Methods

C-terminal phage vector display

The DNA oligonucleotides used for phage engineering are shown in Supplementary Table A.2. Using the Quik Lightning Site-Directed Mutagenesis Kit (Stratagene), silent mutations were inserted into the 983 vector (Ghosh, D. *et al.*, unpublished) at the C-termini of pIII, pVI, and pIX. The starting vector contained the biotin acceptor peptide (GLQDIFEAQKIEWHE) fused to the N-terminus of pIX, DSPHTELP on N-terminus of pVIII and SPARC (secreted protein, acidic and rich in cysteine) binding peptide (SPPTGINGGG) on the N-terminus of pIII. For the pIII, cut sites for *Bsp*EI and *Bcl*I were introduced using pIIIBspEIBclITop and pIIIBspEIBclIBottom oligonucleotides. For pVI, an *Aat*II cut site was introduced using pVIAatIITop and pVIAatIIBottom oligonucleotides and the *Age*I cut site was introduced using the oligonucleotide pair pVIAgeITop and pVIAgeIBottom. For pIX, a *Bsp*HI site existed previously in the genome and a *Spe*I cut site was introduced using the pIXSpeITop and pIXSpeIBottom oligonucleotide pair.

For pIII C-terminal display, the vector was digested with *Bsp*EI and *Bcl*I. Three pairs of annealed oligos: pIIILPETGGHA-C and pIIILPETGGHA-NC, pIII 1GLPETGGHA-C and pIII1GLPETGGHA-NC, and pIII3GLPETGGHA-C and pIII 3GLPETGGHA-NC were annealed and inserted into the digested vector. To determine whether the insert had been incorporated into the phage genome, PCR reactions were performed on the ligations using the coding primer from the inser-

tion (pIIILPETGGHA-C, pIII1GLPETGGHA-C, and pIII3GLPETGGHA-C) and pVISeq as primers. The PCR reactions were analyzed by agarose gel stained with SYBR Safe Stain (Life Technologies) and visualized using a Gel Doc 2000 Gel Documentation System (BioRad). The same procedure was performed for pVI and pIX C-terminal insertions using their respective enzymes and insertion oligos. pVI PCR reactions used the non-coding strand from the insertion with pIIIBspEIBclITop as primers. The pIX PCR reactions used the coding strand of the insertion and -96gIII as primers.

The ligation reactions were transformed into the XL-1 Blue bacterial strain, plated in agar top on LB agar plates containing 1mM IPTG, 40μg/mL X-Gal, and 30μg/mL tetracycline. Plaques were selected and DNA was isolated and sequenced to check for the insertion.

Construction and amplification of LoopXa-pIII phage

The oligonucleotides used in engineering the LoopXa-pIII phage are shown in Supplementary Table A.2. Phage containing the loop with a trypsin cut site was constructed from an M13KE vector (New England Biolabs). The vector was digested with *Acc*65I and *Eag*I. The annealed oligonucleotides pIIILoop-C and pIIILoop-NC were annealed and ligated into the digested vector. The Loop-pIII phage was converted to LoopXa-pIII phage by mutagenesis using the Quik II Site-Directed Mutagenesis kit (Stratagene) with oligonucleotides pIIILoopXaTop and pIIILoopXaBottom. The phage vectors were plated as described in C-terminal phage vector display section.

For LoopXa-pIII phage amplification, the *E. coli* strain ER2738 (New England Biolabs) in LB media supplemented with 30μg/mL tetracycline, was infected with phage for at least 12hrs at 37°C. The cultures were centrifuged at 12000g for 20min to remove bacteria and the phage was precipitated overnight from the supernatant at 4°C with the addition of 1/5 of the supernatant volume of 20% PEG8000/2.5M

NaCl solution. Upon centrifugation at 13500g for 20min, the pellet was resuspended in 25mM Tris, 150mM NaCl, pH. 7.0-7.4 (TBS). For further purification, this resuspension was subjected to two rounds of centrifugation/precipitation. The final phage concentration averaged between 10^{13}-10^{14} plaque forming units (pfu) per mL as determined by UV-vis spectrometry [76].

Sortase-mediated reactions

Sortase reactions were performed as indicated in the figures. A typical sortase reaction for labeling LoopXa-pIII phage involved 160nM phage, 30μM $SrtA_{aureus}$, 230nM Factor Xa, 100μM nucleophile incubated for 5hrs at room temperature. The concentration reported for G_3-CtxB is the monomer concentration. The sortase labeling reactions with GGGK(TAMRA) were monitored by SDS-PAGE under reducing and non-reducing conditions followed by fluorescent imaging and staining with Coomassie. For the CtxB labeling reactions, the reactions were analyzed by SDS-PAGE in reducing and non-reducing conditions followed by immunoblot using an anti-pIII antibody (New England Biolabs) and anti-CtxB (GenWay Biotech).

For the end-to-end structures, 120nM LoopXa-pIII phage, 180nM G_5-pIII phage, 230nM Factor Xa, 30μM $SrtA_{aureus}$, and 10mM $CaCl_2$ in TBS incubated for 60hrs at room temperature. The reaction was diluted in TBS such that the LoopXa-pIII concentration was below 10nM and purified by PEG8000/NaCl precipitation. After removing the supernatant, the reaction was resuspended in water.

Atomic Force Microscopy (AFM)

Phage preparations were diluted in water to a concentration of 2x10^{11} pfu/mL. 100μL of the phage solution was deposited on a freshly cleaved mica disc. AFM images were taken on a Nanoscope IV (Digital Instruments) in air using tapping mode. The tips had spring constants of 20-100N/m driven near their resonant frequency of 200-

400kHz (MikroMasch). Scan rates were approximately 1Hz. Images were leveled using a first-order plane fit to remove sample tilt.

Dynamic Light Scattering (DLS)

DLS measurements were obtained with a DynaPro NanoStar (Wyatt Technology). Phage mixtures were diluted to $\sim 10^{12}$ pfu/mL in water. Samples under reducing conditions contained 0.1M DTT. 8μL of the samples was loaded into a microcuvette. Samples from each experiment were measured 25 times and the results were averaged by cumulant analysis. Autocorrelation functions were used as a direct comparison of linkages because multi-phage structures have a slower Brownian motion causing the signal correlation to be delayed to longer relaxation times.

Miscellaneous

Expression and purification of $\text{SrtA}_{\text{aureus}}$ and G_3-CtxB were performed as described [47]. The sortase reactions were analyzed on 10% Laemlli SDS-PAGE gels. The GGGK(TAMRA) peptide was obtained from the Swanson Biotechnology Center. For mass spectrometry, the protein bands of interest were excised, subjected to protease digestion, and analyzed by electrospray ionization tandem mass spectrometry (MS/MS). Fluorescent gel images were obtained using a variable mode imager (Typhoon 9200; GE Healthcare).

Chapter 4

Formation of phage networks using DNA hybridization

4.1 Introduction

The ordering of materials at the nanoscale level is a significant challenge. The properties of the material and their interactions with other materials are sensitive to their patterning and orientation. This patterning is important for the construction of electronic [100], optical [101, 102], and plasmonic devices [103, 104]. Current high throughput lithography techniques can generate features at the tens of nanometers scale (∼30nm) [105], but it is difficult to generate these patterns over large areas [106]. To achieve the desired organization of material on larger areas, a self assembly approach is necessary to generate such high resolution organization.

DNA hybridization is a crosslinking method that can self assemble these nanoscale features and has been used to produce nanostructures [13]. Unpaired single stranded DNA (ssDNA) functions as a flexible linker to control the spacing between the materials. The specific hybridizing sequence between strands can be tuned to impart specificity and alter the melting temperature. These connections also provide versatility as DNA hybridization is reversible under heating and cooling. DNA crosslinking has been used to order spherical nanoparticles [107, 108] and icosahedral viruses [109] into crystallized lattices, while changes to the hybridizing sequence alter the packing of the particles [110, 111].

M13 bacteriophage is a material scaffold, which has been used as a scaffold

for the nucleation of materials to form nanowires [28, 35, 36]. It has a cylindrical shape 880nm long and 6nm in diameter which creates a high surface area to volume ratio, a desirable trait for nucleating materials. The phage is composed of five capsid proteins with 5 copies of pIII and pVI located at one end and 5 copies of pVII and pIX located at the other. The body is composed of 2700 copies of pVIII. These five M13 capsid proteins can be genetically modified through a process known as phage display [53, 54]. Using phage display of peptide libraries, peptides have been identified that when expressed on phage can bind materials including plasmonic materials like gold and silver [28, 35] as well as carbon nanotubes [30, 38].

The multiple capsid proteins of phage provide three distinct motifs for attachment: the body and both ends. By controllably connecting the phage, materials can be connected at the nanoscale. Previous work to create ordered phage used high concentrations of phage to form liquid crystals [98], which can be crosslinked to form fibers [112]. When phage is at lower concentrations, phage form unordered but porous hydrogels (Burpo, F. J. *et al.*, unpublished). In the formation of fibers and hydrogels, gluteraldehyde is used to crosslink the exposed primary amines of the phage capsid proteins. This crosslinking method, however, is permanent and not specific to a particular capsid protein and any functionality that has been added to the pVIII or other capsid proteins may be hindered or eliminated. In these structures, there is no separation between the crosslinked phage. DNA hybridization offers advantages over these crosslinking chemistries with specificity, reversibility, and flexibility in the connections. Combining these materials, we can build DNA crosslinked phage networks.

In order to build such networks, a conjugation chemistry specific to a single capsid protein is required to attach DNA. In our network, we aimed to connect the pVIII of two phages to allow for possible side-by-side alignment of the phage (Fig. 4.1). The connection must be capsid protein specific as other connections can cause

irregular spacing as a result of disoriented phage. Therefore, chemistries not specific to a single capsid protein, such as EDC, are not ideal for attaching the DNA to phage to build these structures. Although the pVIII has been site-specifically biotinylated [52], streptavidin-biotin linkages are not ideal for attaching DNA since the streptavidin tetramer is denatured upon heating, which limits the recyclability of the linkages.

Sortase A from *Streptococcus pyogenes* ($SrtA_{pyogenes}$) has been used to label a pVIII protein with an N-terminal double alanine motif [52]. The enzyme recognizes an LPXTA peptide motif and cleaves after the threonine forming an acyl-intermediate that is resolved by nucelophilic attack by an N-terminal polyalanine motif [45]. Using this enzymatic labeling technique and a DNA peptide conjugate, we can covalently attach DNA to the pVIII protein.

We crosslink DNA labeled phage with base pair hybridization and demonstrate that these structures can be dispersed by heat and reformed upon cooling below the calculated melting temperature of the crosslinking oligonucleotides. This initial demonstration that DNA crosslinked phage structures are possible lays the groundwork for creating recyclable phage networks and combining phage and DNA nanotechnologies.

4.2 Results

4.2.1 Attaching DNA to pVIII using sortase-mediated reactions

To label phage with DNA, we created a substrate for $SrtA_{pyogenes}$ that contains both DNA and the LPETAA substrate motif. We selected DNA oligonucleotides (DNA A and DNA B), shown in Table 4.1, which had been used to order gold nanoparticles [107]. The oligonucleotides contain 15bp complementary regions at their 5' ends.

At the 3' end, there is a 35bp ssDNA linker, whose length was found to be critical for achieving crystallization of the nanoparticles [107]. A thiol group was installed at the 3' end for chemical conjugation. We attached the DNA to an L(maleimide)-PETAA peptide and purified the reaction by size exclusion chromatography. We monitored the conjugation reaction by MALDI-TOF mass spectrometry (Fig. 4.2). For both DNA oligonucleotides, we observed a 700Da shift in molecular weight when the peptide is reacted with the DNA. This product is consistent with the size of the DNA-LPETAA fusion. The conjugation reaction went to completion as the peak corresponding to unreacted DNA is not observed after the reaction with the peptide. No peak was observed at 700Da corresponding to free peptide (data not shown).

Table 4.1: **Oligonucleotides for crosslinking phage**

Name	**Sequence (5'-3')**
DNA A	TACTTCCAATCCAATTCTTGTGTCGATAGGTCGGTTGCTTTTTTTTTTTT
DNA B	ATTGGATTGGAAGTATCTTGTGTCGATAGGTCGGTTGCTTTTTTTTTTTT

$SrtA_{pyogenes}$ requires its nucelophile to have an N-terminal double alanine sequence [45]. Therefore, the pVIII, which includes an N-terminal alanine that cannot be removed, was modified to display two N-terminal alanines (A_2-pVIII phage). This phage was incubated with $SrtA_{pyogenes}$, and DNA A-LPETAA or DNA B-LPETAA. We purified the reaction using PEG8000/NaCl precipitation to remove excess DNA-peptide and $SrtA_{pyogenes}$. Removing free DNA-peptide is necessary as it can hybridize with DNA on the surface of phage labeled with the complementary strand, and thus block sites for crosslinking.

We analyzed the reactions under reducing conditions on an SDS-PAGE gel followed by staining with SYBR Gold, a DNA intercalating dye (Fig. 4.3). When the DNA-peptide, phage, and $SrtA_{pyogenes}$ were incubated at 37°C, we detected a 25kDa DNA containing species consistent with the size of the DNA-pVIII. When the gel was subsequently stained with Coomassie (Fig. 4.3), a 25kDa polypeptide was observed

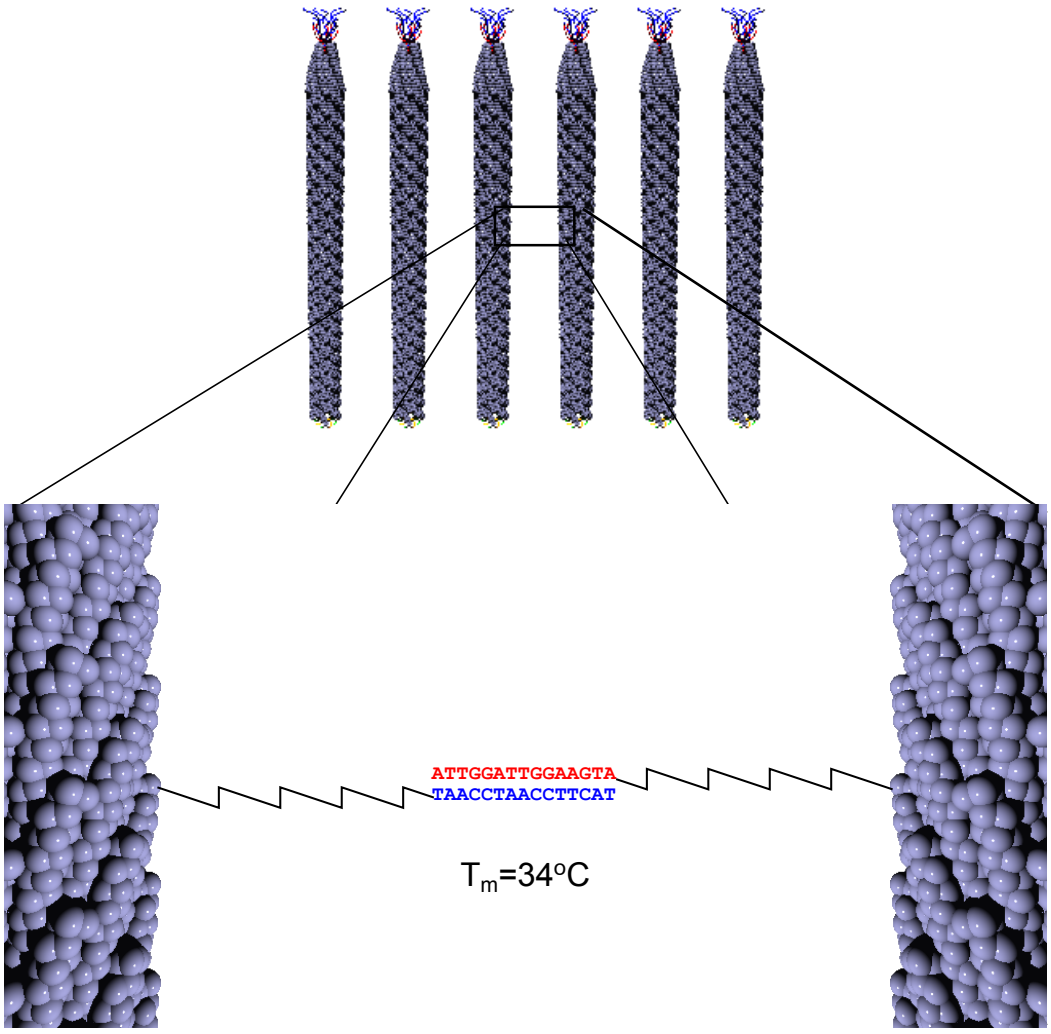


Figure 4.1: **Schematic of phage crosslinking by DNA hybridization**. The pVIII proteins of different phage are labeled complementary strands of ssDNA. Hybridization of the DNA strands form linkages between the body of two phage.

that coincides with the SYBR Gold reactive species supporting the presence of a DNA protein conjugate. We confirmed by mass spectrometry that this species contained part of the pVIII protein for reaction for both DNA A and DNA B (Fig. 4.4). In the samples that were purified by PEG/NaCl precipitation, we did not detect the free DNA-peptide and $SrtA_{pyogenes}$-DNA intermediate showing our purification was successful. Although the pVIII protein showed staining, this staining was not specific and present even when DNA-peptide conjugate was not present.

4.2.2 Crosslinking phage using DNA hybridization

Although the calculated melting temperature of the oligonucleotides is 34°C, the avidity of having multiple copies of the DNA may increase the melting temperature between the crosslinked phages making it impossible to break up the phage aggregates without going to high temperatures. To produce phage which would have less avidity effects in its crosslinking, we performed a second labeling reaction with 5-fold lower concentration of the DNA-peptide. The reaction was purified by PEG8000/NaCl purification and analyzed the reactions by SDS-PAGE and stained the gel with SYBR Gold (data not shown). Compared with our initial labeling, we observed less DNA-pVIII for both DNA A and DNA B labeling.

In order to crosslink the phage, the labeled phage was subjected to the annealing process used to hybridize complementary DNA strands. Phage labeled with DNA A (Phage A) and phage labeled with DNA B (Phage B) as well as the phage with less DNA labeling (Phage A-less and Phage B-less) were mixed in a 1:1 ratio and heated to 95°C and cooled to room temperature. We positively stained the samples and visualized them by transmission electron microscopy (TEM) (Fig. 4.5a). When Phage A and Phage B (Phage AB) were annealed, we observed crosslinking of the phage and the formation of networks comprised of overlapping phage. When the Phage A-less and Phage B-less were annealed (Phage AB-less), phage connections

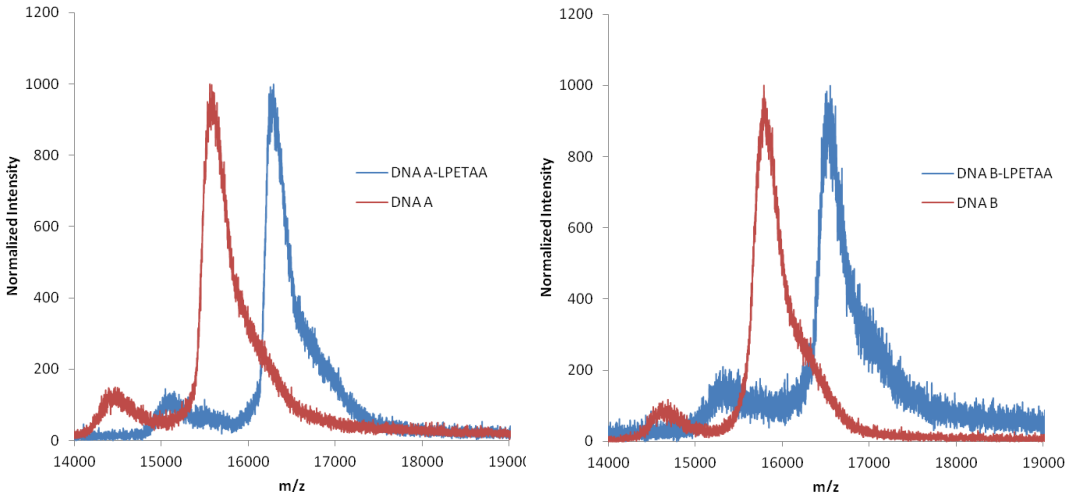


Figure 4.2: **MALDI-TOF mass spectrometry of DNA-peptide conjugate**. Thiolated DNA A and DNA B were incubated with L(maleimide)-PETAA to form DNA-peptides. The reactions were monitored by MALDI-TOF mass spectrometry.

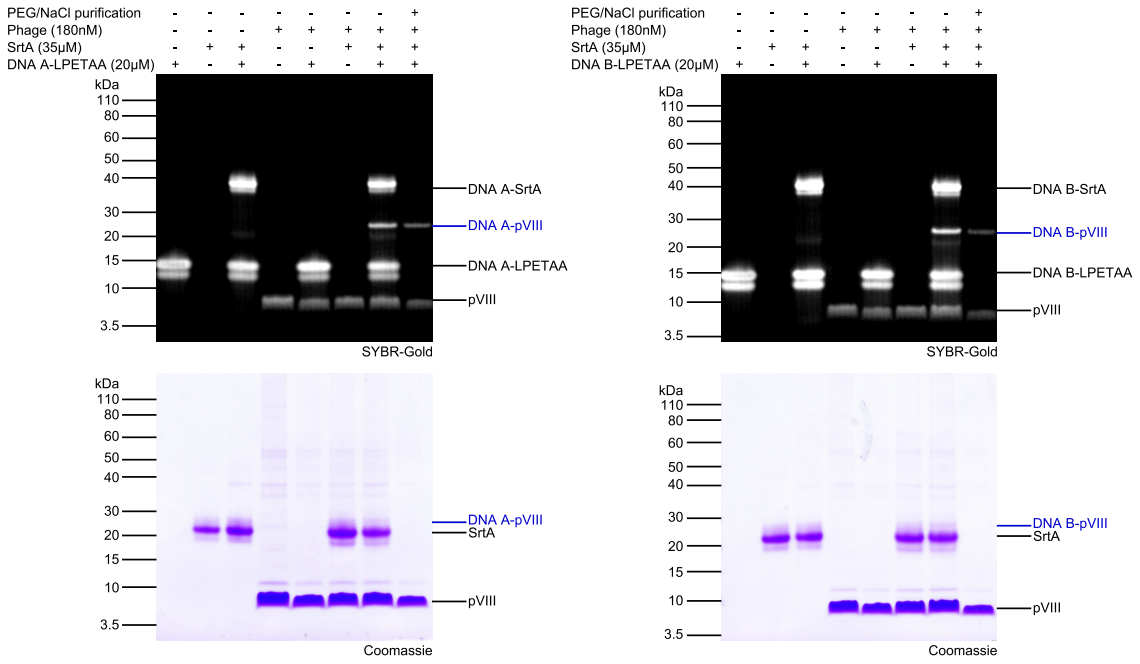


Figure 4.3: **Labeling of pVIII with DNA**. A_2-pVIII phage was incubated with $SrtA_{pyogenes}$ and DNA A-LPETAA (left panel) and DNA B-LPETAA (right panel) at 37C for 2hrs. The reactions were monitored by SDS-PAGE under reducing conditions. The gel was stained with SYBR Gold (top panels) and Coomassie (bottom panels). The molecular weight markers are indicated on the left.

were seen, but the network of phage was not as dense (Fig. 4.5a). We did not observe phage networks by TEM when unlabeled phage, Phage A, and Phage B were heated and cooled (Fig. 4.5a).

We analyzed the phage networks using dynamic light scattering (DLS) to monitor the size of these phage networks (Fig. 4.5b). The crosslinked structures had an increase in hydrodynamic radius when compared to individual phage (Table 4.2). When the phage were annealed together, we saw a population in the hydrodynamic radius intensity distribution with a radius exceeding 10000nm representing the existence of large crosslinked structures (Fig. 4.5b). We did not observe this population in the intensity distributions of the unlabeled phage, Phage A, and Phage B.

Table 4.2: **Hydrodynamic radius of phage networks**

Sample	**Hydrodynamic radius (nm)**
Unlabeled Phage	336.345
Phage A	348.408
Phage B	367.029
Phage AB	602.151
Phage AB-less	569.332

4.2.3 Recyclability of phage crosslinking

An advantage of using DNA crosslinking is that the process is recyclable when the structures are heated and cooled. To demonstrate this feature, we heated and cooled the phage networks and monitored their structures by DLS (Fig. 4.6a). The samples were heated to 90°C before being cooled slowly to 20°C. During this cycle, we measured the hydrodynamic radius (Table **??**) and intensity distribution at the temperatures indicated (Fig. 4.6). Upon heating both Phage AB and Phage AB-less above 80°C, the population with hydrodynamic radii greater than 10000nm disappeared representing the dispersal of the phage networks. The intensity distribution

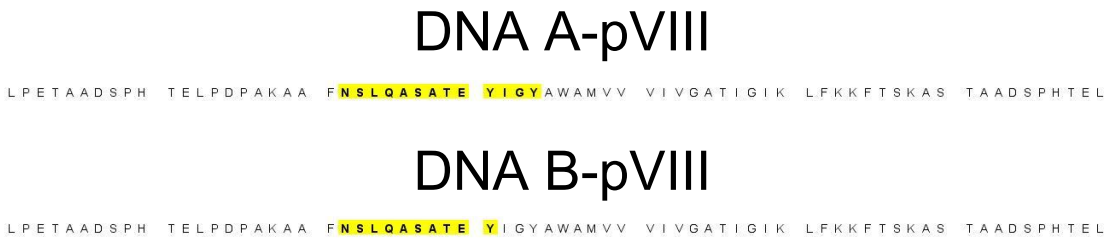


Figure 4.4: **Characterization of DNA-pVIII fusion proteins by mass spectrometry**. The DNA-polypeptide corresponding to DNA A-pVIII and DNA B-pVIII was excised from the SDS-PAGE gel and digested with chymotrypsin. The resulting peptides were analyzed by liquid chromatography MS/MS. Peptides positively identified by sequence are highlighted in yellow.

and hydrodynamic radius were consistent with phage that was not crosslinked.

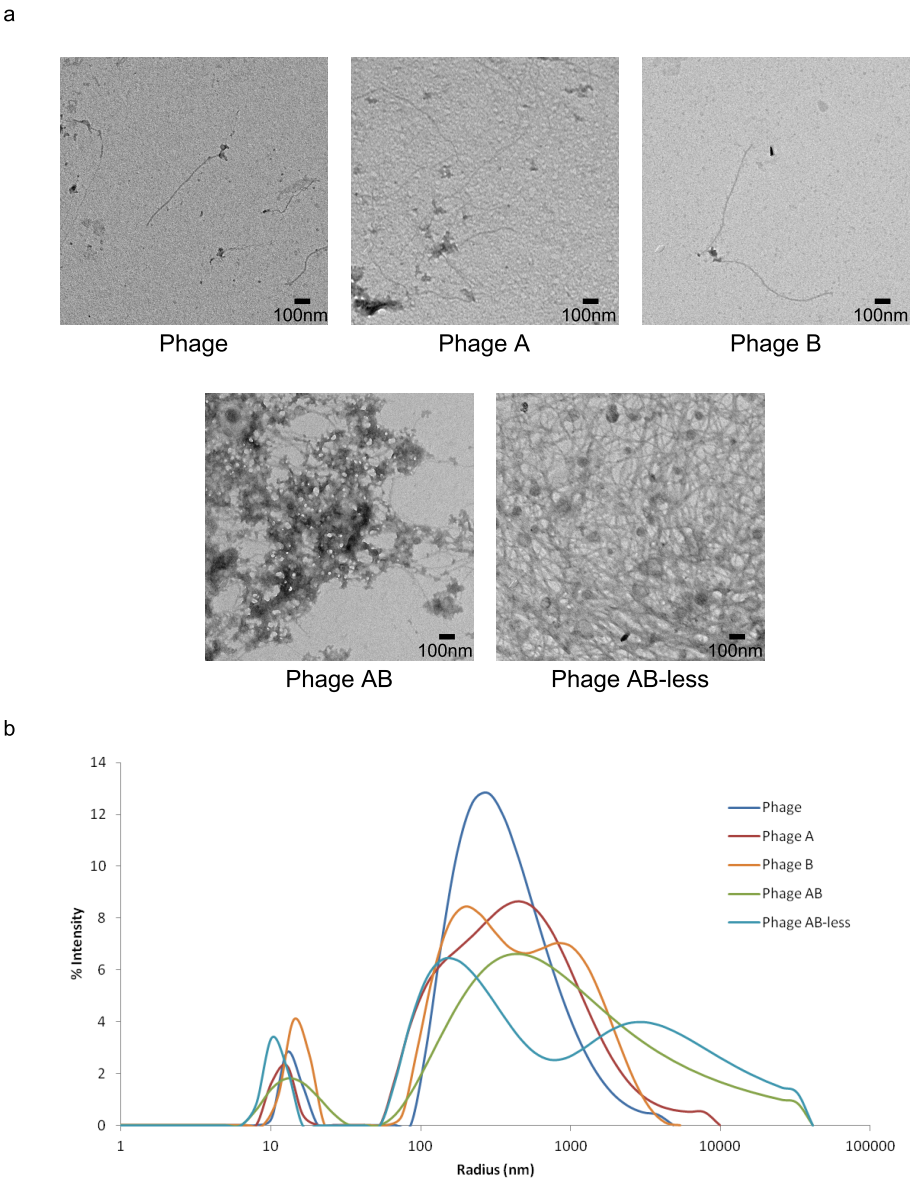


Figure 4.5: **DNA hybridized phage networks**. Phage A and Phage B as well as Phage A-less and Phage B-less were mixed and annealed (see Methods section for details) to form phage networks. The samples were analyzed by transmission electron microscopy (a) and dynamic light scattering (b).

Upon cooling to 20°C, we observed that the population with hydrodynamic radii greater than 10000nm returned (Fig. 4.6a) and the hydrodynamic radius increases (Table **??**). For Phage AB, we saw the large radii population returning at 32.0°C. For Phage AB-less, crosslinking returned at 21.4°C. This observation supports that the number of DNA displayed on the phage affects the annealing temperature of the crosslinking. When unlabeled phage, Phage A, and Phage B were heated and cooled, we did not observe any changes in the hydrodynamic radius or intensity distribution (Fig. 4.7).

After cooling, we collected Phage AB and Phage AB-less and visualized the samples by TEM after positive staining (Fig. 4.6b). We saw phage networks in both samples (Fig. 4.6b) demonstrating the recyclability of the phage crosslinking by DNA hybridization. We noticed less dense crosslinking than that found in Phage AB and Phage AB-less before DLS analysis. We accounted for this decrease as the samples were diluted before DLS analysis to not saturate the signal, and the second annealing cycle was performed at a lower phage concentration than the initial annealing.

4.3 Discussion

Using phage displaying complementary DNA strands, we created networks of crosslinked phage which we monitored by TEM and DLS. These networks were dispersed by heat and reformed upon cooling. The reformation of the networks did not occur until the temperature had been lowered beneath the calculated melting temperature of the hybridizing oligonucleotides (T_m=34°C), which supports that hybridization is involved in these interactions. Furthermore, we performed the crosslinking using phage displaying different amounts of DNA. The crosslinking returned at a higher temperature for the phage with more copies of the DNA. These observations demonstrate that annealing properties can be altered by the number of DNA

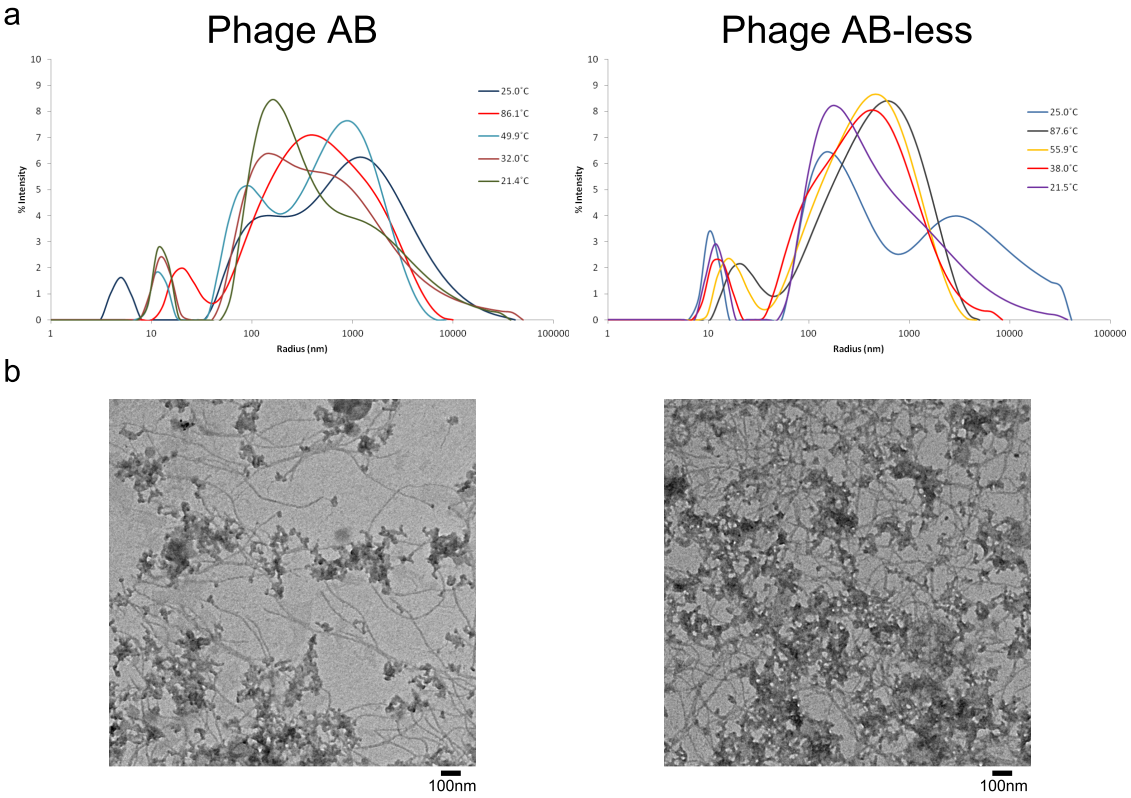


Figure 4.6: **Recyclability of phage networks**. Phage AB and Phage AB-less, which formed networks, were heated and cooled. The samples were monitored by dynamic light scattering (a) and measurements were made at the temperatures indicated. After cooling, the samples were positively stained and visualized by transmission electron microscopy (b).

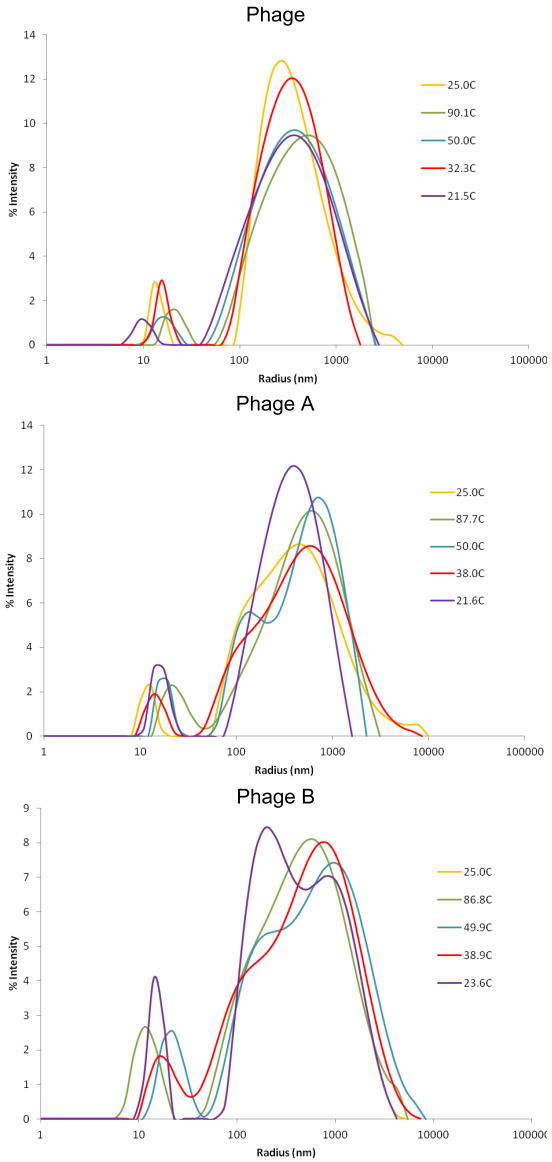


Figure 4.7: **Heating and cooling of phage labeled with DNA**. Phage, Phage A, and Phage B were heated and cooled. The samples were monitored by dynamic light scattering and measurements were made at the temperatures indicated.

molecules displayed on phage.

These experiments serve as proof that we can use hybridized DNA to produce multi-phage structures, which may be used for the deposition of materials. The phage networks formed in this work require optimization to obtain crystalline patterning of the phage, but this is an initial step to construct an ordered array of nanorods. This ordering may be achieved by a combination of optimizing the melting temperature of the hybridized DNA, increasing the phage concentration to the point where phage can form ordered liquid crystals [98, 112], or using a pre-patterned scaffold to align DNA labeled phage and fill in the spaces with phage displaying complementary DNA to create patterned and aligned phage.

These aligned phage materials have specific applications in the construction of plasmonic devices with optical and biosensor applications [113–115]. Plasmons are the product of free electron oscillations in a conducting material. When these electrons are subjected to an external electric field such as that from incident light, electrons shift in the material to cancel out the electric field. Once the field is removed, the electrons begin oscillating at the plasma frequency of the material. This phenomenon has optical implications as light below the plasma frequency are reflected, while light above the frequency is transmitted. These optical properties are highly sensitive to the dielectric environment, which makes these materials valuable sensors of binding events that alter the dielectric constant. In addition to biosensing, the absorption properties of these materials can be used to construct thinner solar cells [116]. In these devices, spacing is critical as if the particles were to be in contact, the electron oscillations would flow between the materials mitigating the plasmonic effects.

Without ordering, these recyclable networks have applications as filtration meshes. Using phage display, the pVIII of the phage can be engineered with the ability to bind and trap contaminants. The length of the crosslinking DNA can alter the melting temperature as well as the size of the pores in the mesh by changing the sepa-

ration between linked phage. Using heat, the meshes are dispersed and contaminants are released with temporary denaturing of the pVIII by temperature increase. Cooling of these phages would reform the mesh with the filtered contaminants removed.

The crosslinking of phage using DNA hybridization can be extended to other phage to phage linkages including self-assembly of multi-phage nanorods composed of any number of different phages. Different DNA sequences provide an endless supply of orthogonal chemistries which can be used to construct multimer connections in a specific order as programmed into the DNA sequences, which has been used to build molecular wires [117]. This specificity allows the multi-phage structures to be built in one reaction instead of building the structures step by step. The specific hybridizing sequences can be designed with different melting temperatures making it possible to break the structure sequentially by heating. It may also be possible to build multi-way junctions between phage using more complex DNA interactions such as Holliday junctions. This can even be extended into the construction of DNA origami structures spanning the space between phage. These DNA higher order structures can increase the rigidity of the linkages or add additional functionality.

DNA has been used in a wide range of technologies including biosensor and material applications [12, 118]. Using sortase-mediated reactions, we can combine these technologies with those already developed on phage. In the merging of DNA and phage nanotechnologies, phage provides a stiffer scaffold given its persistence length is greater than that of double stranded DNA [119]. To construct more complex devices, multiple functionalities may need to be placed on a single material scaffold, which is addressed by the multiple capsid proteins on M13 phage. The protein scaffold of phage provides a variety of amino acids greater than that of oligonucleotides, supplying more targets for chemical modification.

Combining the two biomaterials, a single phage can be designed with two functionalities using DNA aptamers as the addition of the DNA to pVIII does not

disrupt functionality engineered into the N-terminus of the pVIII protein. Instead of the hybridizing nucleotides that we added, a DNA aptamer selected for binding properties can be attached to give a second binding motif. This DNA may make the phage capable of binding two moieties that require each other for full function. Given short DNA structures have been used as switches induced by chemical or pH change [14, 120], it may be possible to have the second binding controlled by environmental factors.

This work serves as a proof-of-concept that DNA can be displayed on the surface of phage in a site-specific manner, and that the DNA maintains its hybridizing properties in the formation of networks with recyclable connections. We view these linkages as initial steps to building higher ordered structures for plasmonic applications, and we believe this labeling strategy provides a clear bridge between DNA and phage nanotechnology.

4.4 Methods

DNA-peptide conjugation

The crosslinking DNA, shown in Table 4.1, contained a disulfide group at their 3' ends. The thiol group on the DNA was activated overnight in 0.1M DTT in PBS at 37°C. The DNA was then purified from excess DTT on a NAP5 column (GE Healthcare) and eluted in water. The solution was dried and resuspended in PBS, pH 7.0-7.5. L(maleimide)-PETAA peptide in PBS was added in 2:1 molar excess of the activated DNA and reacted for 5hrs at 37°C. In order to deactivate the excess maleimide, 0.1M DTT was added to the mixture and incubated at 37°C for 15min. The excess DTT was removed by purifying the reaction on a NAP5 column. The DNA peptide was dried under vacuum and resuspended in 25mM Tris, pH 7.0-7.4, and 150mM NaCl (TBS). Before drying by vacuum, samples were collected for

analysis by MALDI-TOF mass spectrometry. The concentration of the DNA-peptide was determined by UV-vis spectrometry using the absorbance at 260nm.

Phage Engineering

A_2-pVIII phage was modified from an M13SK vector [23] that contained a DSPHTELP sequence on its pVIII and biotin acceptor peptide (GLQDIFEAQKIEWHE) on the pIII. The N-terminus of pVIII was modified using AADSPH-pVIII-Top and AADSPH-pVIII-Bottom, shown in Supplementary Table A.1 using the QuikChange II site-directed mutagenesis kit (Stratagene). The generated phage vector was transformed into the XL-1 Blue bacterial strain and plated in agar top on LB agar plates containing 1mM IPTG, 40μg/mL X-Gal, and 30μg/mL tetracycline. Plaques were selected and DNA was isolated and sequenced to check for the modification.

For phage amplification, the *E. coli* strain ER2738 (New England Biolabs) in LB media supplemented with 30μg/mL tetracycline, was infected with phage for at least 12hrs at 37°C. The cultures were centrifuged at 12000g for 20min and the phage was precipitated from the supernatant at 4°C with the addition of 1/5 of the supernatant volume of 20% PEG8000/2.5M NaCl solution. Upon centrifugation at 13500g for 20min, the pellet was resuspended in TBS. For further purification, this resuspension was subjected to two rounds of centrifugation/precipitation. The final phage concentration averaged between 10^{13}-10^{14} plaque forming units (pfu) per mL as determined by UV-vis spectrometry [76].

Sortase-mediated reactions

Sortase reactions were performed as indicated in the figures. A typical sortase reaction includes 180nM A_2-pVIII phage, 30μM $SrtA_{pyogenes}$, 20μM DNA A-LPETAA or DNA B-LPETAA incubated in TBS for 2hrs at 37°C. For Phage A-less and Phage

B-less, the reaction included only 4μM DNA-LPETAA. The reactions were diluted in TBS such that the concentration of DNA-LPETAA was under 600nM. For purification, the phage was precipitated using PEG/NaCl on ice for at least 3hrs. Upon centrifugation at 17000g for 15min, the pellet was resuspended in water. The reactions were analyzed on an SDS-PAGE gel under reducing conditions. The proteins were fixed to the gel followed by rinsing for 3hrs in water before staining with SYBR-Gold (Life Technologies) and Coomassie.

Dynamic Light Scattering (DLS)

In order to form the phage networks, Phage A and Phage B were mixed in a 1:1 ratio in 12μL of water at a concentration of 5x10^{13} pfu/mL. The mixture is heated to 95°C for 2min and cooled to 25°C at a rate of 0.1°C/s. DLS measurements were obtained with a DynaPro NanoStar (Wyatt Technology). These annealed mixtures were diluted to 2x10^{12} pfu/mL in water and 4μL was loaded into a quartz microcuvette. The samples were measured at 25°C before being heated to 90°C for 5min after which the samples were cooled to 20°C at a rate of 6°C/min. Each measurement was made ten times and averaged by cumulant analysis. Intensity distributions of the hydrodynamic radius were used to monitor crosslinking and the presence of objects with an increased radius compared to single phage.

Transmission Electron Microscopy (TEM)

TEM measurements were obtained using a JEOL 200Cx TEM (JEOL, Ltd) operating at 120kV. Samples were diluted in water to a concentration of 1x10^{13} pfu/mL. 5μL of each solution was placed onto a carbon coated 300 mesh copper grid (Electron Microscopy Sciences). After 60s, the remaining liquid was wicked off and the sample was stained with 0.5% uranyl acetate (Electron Microscopy Sciences) in water. Due to the high affinity of phage for metal ions, the samples were positively stained.

Miscellaneous

Expression and purification of $SrtA_{pyogenes}$ were performed as described [47]. The sortase reactions were analyzed on 4-12% Bis-Tris SDS-PAGE gels (Life Technologies) with MES running buffer. Fluorescent gel images were collected on an AlphaImager 2200 (Alpha Innotech). MALDI-TOF mass spectrometry of DNA-peptide conjugates was performed by the Swanson Biotechnology Center. The L(maleimide)-PETAA peptide was obtained from the Swanson Biotechnology Center.

Appendix A

Supplementary Material

Protein Coverage:

Sequence	MH+	% Mass	AA	% AA
MVSKGEELFTGVVPILVELDGDVNGHK	2882.50	4.10	1 - 27	4.12
GEELFTGVVPILVELDGDVNGHK	2437.26	3.47	5 - 27	3.51
FSVSGEGEGDATYGK	1503.66	2.14	28 - 42	2.29
FSVSGEGEGDATYGKLTLK	1958.97	2.79	28 - 46	2.90
FICTTGK	769.39	1.10	47 - 53	1.07
LPVPWPTLVTTLTYGVQCFSRYPDHMK	3149.60	4.48	54 - 80	4.12
YPDHMKQHDFFK	1592.73	2.27	75 - 86	1.83
YPDHMKQHDFFKSAMPEGYVQER	2840.29	4.04	75 - 97	3.51
QHDFFK	821.39	1.17	81 - 86	0.92
QHDFFKSAMPEGYVQER	2068.95	2.95	81 - 97	2.60
QHDFFKSAMPEGYVQERTIFFK	2705.32	3.85	81 - 102	3.36
SAMPEGYVQER	1266.58	1.80	87 - 97	1.68
SAMPEGYVQERTIFFK	1902.94	2.71	87 - 102	2.44
SAMPEGYVQERTIFFKDDGNYK	2595.22	3.69	87 - 108	3.36
TIFFKDDGNYK	1347.66	1.92	98 - 108	1.68
TIFFKDDGNYKTR	1604.81	2.28	98 - 110	1.98
DDGNYKTRAEVK	1395.69	1.99	103 - 114	1.83
TRAEVKFEGDTLVNR	1734.91	2.47	109 - 123	2.29
AEVKFEGDTLVNR	1477.76	2.10	111 - 123	1.98
FEGDTLVNR	1050.52	1.50	115 - 123	1.37
FEGDTLVNRIELK	1533.83	2.18	115 - 127	1.98
IELKGIDFK	1062.62	1.51	124 - 132	1.37
IELKGIDFKEDGNILGHK	2026.10	2.88	124 - 141	2.75
GIDFKEDGNILGHK	1542.79	2.20	128 - 141	2.14
EDGNILGHK	982.50	1.40	133 - 141	1.37
EDGNILGHKLEYNYNSHNVYIMADK	2937.38	4.18	133 - 157	3.82
LEYNYNSHNVYIMADK	1973.91	2.81	142 - 157	2.44
LEYNYNSHNVYIMADKQK	2230.06	3.17	142 - 159	2.75
LEYNYNSHNVYIMADKQKNGIK	2642.30	3.76	142 - 163	3.36
QKNGIK	687.41	0.98	158 - 163	0.92
NGIKVNFK	919.54	1.31	160 - 167	1.22
NGIKVNFKIR	1188.72	1.69	160 - 169	1.53
RDHMVLLEFVTAAGITLGMDELYK	2722.39	3.88	216 - 239	3.66
DHMVLLEFVTAAGITLGMDELYK	2566.29	3.65	217 - 239	3.51
LPETGGGGGSAETVESCLAK	1862.88	2.65	240 - 259	3.05
SHTENSFTNVWK	1449.68	2.06	260 - 271	1.83
SHTENSFTNVWKDDK	1807.82	2.57	260 - 274	2.29
DDKTLDR	862.43	1.23	272 - 278	1.07
FRNRQGALTVYTGTVTQGTDPVK	2509.32	3.57	390 - 412	3.51
QGALTVYTGTVTQGTDPVK	1936.00	2.76	394 - 412	2.90
TYYQYTPVSSK	1336.64	1.90	413 - 423	1.68
TYYQYTPVSSKAMYDAYWNGKFR	2839.32	4.04	413 - 435	3.51
AMYDAYWNGK	1218.52	1.73	424 - 433	1.53
MANANK	648.31	0.92	512 - 517	0.92
MANANKGAMTENADENALQSDAKGK	2579.18	3.67	512 - 536	3.82
GAMTENADENALQSDAK	1764.77	2.51	518 - 534	2.60
GVFAFLLYVATFMYVFSTFANILR	2790.47	3.97	628 - 651	3.66
GVFAFLLYVATFMYVFSTFANILRNK	3032.61	4.32	628 - 653	3.97
Totals:	21693.28	30.88	199	30.38

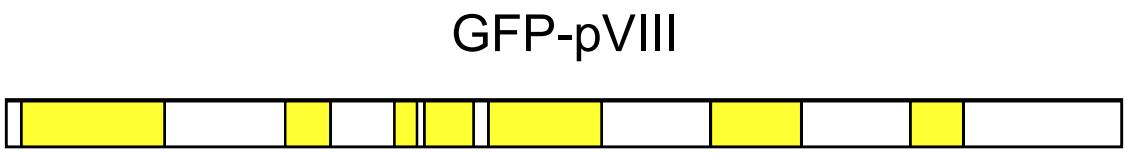
Figure A.1: **Characterization of the GFP-pIII conjugate by mass spectrometry**. The polypeptide corresponding to GFP-pIII was excised from the SDS-PAGE gel and digested with trypsin. The resulting peptides were analyzed by liquid chromatography MS/MS. Peptides positively identified by sequence are highlighted in yellow and bold.

GFP-pIX

Protein Coverage:

Sequence	MH+	% Mass	AA	% AA
MVSKGEELFTGVVPILVEL	2060.13	6.24	1 - 19	6.42
MVSKGEELFTGVVPILVELDG	2232.18	6.76	1 - 21	7.09
ELFTGVVPILVEL	1428.83	4.33	7 - 19	4.39
DGDVNGHKFSVSGEGEG	1690.73	5.12	20 - 36	5.74
DVNGHKFSVSGEGEG	1518.68	4.60	22 - 36	5.07
EGEGDATYGKLTLKFI	1741.90	5.28	33 - 48	5.41
CFSRYPDHMKQHDFFKSAMP	2472.10	7.49	71 - 90	6.76
DHMKQHDFFKSAMP	1718.78	5.21	77 - 90	4.73
DFFKSAMPEGYVQ	1518.69	4.60	83 - 95	4.39
DFFKSAMPEGYVQERTIFFK	2440.20	7.39	83 - 102	6.76
EGYVQERTIFFK	1516.78	4.60	91 - 102	4.05
ERTIFFK	940.53	2.85	96 - 102	2.36
ERTIFFKDDGNYKTRA	1960.99	5.94	96 - 111	5.41
DDGNYKTRA	1039.48	3.15	103 - 111	3.04
DDGNYKTRAEVKFEG	1728.82	5.24	103 - 117	5.07
DGNYKTRA	924.45	2.80	104 - 111	2.70
EVKFEG	708.36	2.15	112 - 117	2.03
EGDTLVNRI	1016.54	3.08	116 - 124	3.04
EGDTLVNRIELKGI	1556.86	4.72	116 - 129	4.73
DTLVNRI	830.47	2.52	118 - 124	2.36
DTLVNRIELKGI	1370.80	4.15	118 - 129	4.05
DTLVNRIELKGIDFK	1760.99	5.33	118 - 132	5.07
DTLVNRIELKGIDFKE	1890.03	5.73	118 - 133	5.41
ELKGIDFK	949.54	2.88	125 - 132	2.70
EDGNILGHKL	1095.58	3.32	133 - 142	3.38
DGNILGHKL	966.54	2.93	134 - 142	3.04
DGNILGHKLEYNYNSHNVYIMA	2565.22	7.77	134 - 155	7.43
EYNYNSHNVYIMA	1617.70	4.90	143 - 155	4.39
DKQKNGIKVNFKIRHNI	2052.18	6.22	156 - 172	5.74
DKQKNGIKVNFKIRHNIE	2181.23	6.61	156 - 173	6.08
DGSVQLA	689.35	2.09	174 - 180	2.36
DGSVQLADHYQQNTPIG	1842.86	5.58	174 - 190	5.74
DHYQQNTPIG	1172.53	3.55	181 - 190	3.38
DHYQQNTPIGDGPVLLP	1863.92	5.65	181 - 197	5.74
DGPVLLP	710.41	2.15	191 - 197	2.36
DGPVLLPDNHYLSTQSALSK	2155.10	6.53	191 - 210	6.76
DGPVLLPDNHYLSTQSALSKDPN	2481.23	7.52	191 - 213	7.77
DNHYLSTQSALSK	1463.71	4.43	198 - 210	4.39
DNHYLSTQSALSKDPNEKR	2203.07	6.67	198 - 216	6.42
DNHYLSTQSALSKDPNEKRDHMVLL	2911.44	8.82	198 - 222	8.45
DPNEKRDHMVLL	1466.74	4.44	211 - 222	4.05
DPNEKRDHMVLLEFVTAAGITLGMD	2772.37	8.40	211 - 235	8.45
DHMVLLEFVTAAGITLGM	1917.98	5.81	217 - 234	6.08
DHMVLLEFVTAAGITLGMD	2033.01	6.16	217 - 235	6.42
EFVTAAGITLGMD	1324.65	4.01	223 - 235	4.39
EFVTAAGITLGMDELYKLP	2068.07	6.27	223 - 241	6.42
DELYKLPETGGGGGYPY	1815.84	5.50	235 - 251	5.74
ETGGGGGYPYDVP	1268.54	3.84	242 - 254	4.39
DVPDYAQGGQGV	1205.54	3.65	252 - 263	4.05
Totals:	17404.73	52.73	155	52.36

Figure A.2: **Characterization of the GFP-pIX conjugate by mass spectrometry**. The polypeptide corresponding to GFP-pIII was excised from the SDS-PAGE gel and digested with AspN. The resulting peptides were analyzed by liquid chromatography MS/MS. Peptides positively identified by sequence are highlighted in yellow and bold.



Protein Coverage:

Sequence	MH+	% Mass	AA	% AA
MVSKGEELFTGVVPILVELDGDVNGH	2754.40	8.46	1 - 26	8.78
GEELFTGVVPILVELDGDVNGHK	2437.26	7.49	5 - 27	7.77
FSVSGEGEGDATYGK	1503.66	4.62	28 - 42	5.07
LPVPWPTLVTTL	1336.79	4.11	54 - 65	4.05
YPDHMK	790.36	2.43	75 - 80	2.03
YPDHMKQHDFFK	1592.73	4.89	75 - 86	4.05
DDGNYK	711.29	2.18	104 - 109	2.03
AEVKFEGDTLVNR	1477.76	4.54	112 - 124	4.39
FEGDTLVNR	1050.52	3.23	116 - 124	3.04
GIDFKEDGNILGHK	1542.79	4.74	129 - 142	4.73
LEYNYNSH	1039.45	3.19	143 - 150	2.70
LEYNYNSHNVYIMADK	1973.91	6.06	143 - 158	5.41
YNSHNVYIMADK	1454.67	4.47	147 - 158	4.05
YNSHNVYIMADKQK	1710.83	5.25	147 - 160	4.73
SHNVYIMADK	1177.57	3.62	149 - 158	3.38
NVYIMADKQKNGIKVN	1834.98	5.64	151 - 166	5.41
VYIMADKQKNGIKVNFK	1996.10	6.13	152 - 168	5.74
TPIGDGPVLLPDNHYLSTQSALSK	2523.31	7.75	188 - 211	8.11
HYLSTQSALSK	1234.64	3.79	201 - 211	3.72
LPETAAGGGGDPAK	1240.62	3.81	241 - 254	4.73
Totals:	14851.19	45.61	137	46.28

Figure A.3: **Characterization of the GFP-pVIII conjugate by mass spectrometry**. The polypeptide corresponding to GFP-pVIII was excised from the SDS-PAGE gel and digested with trypsin. The resulting peptides were analyzed by liquid chromatography MS/MS. Peptides positively identified by sequence are highlighted in yellow and bold.

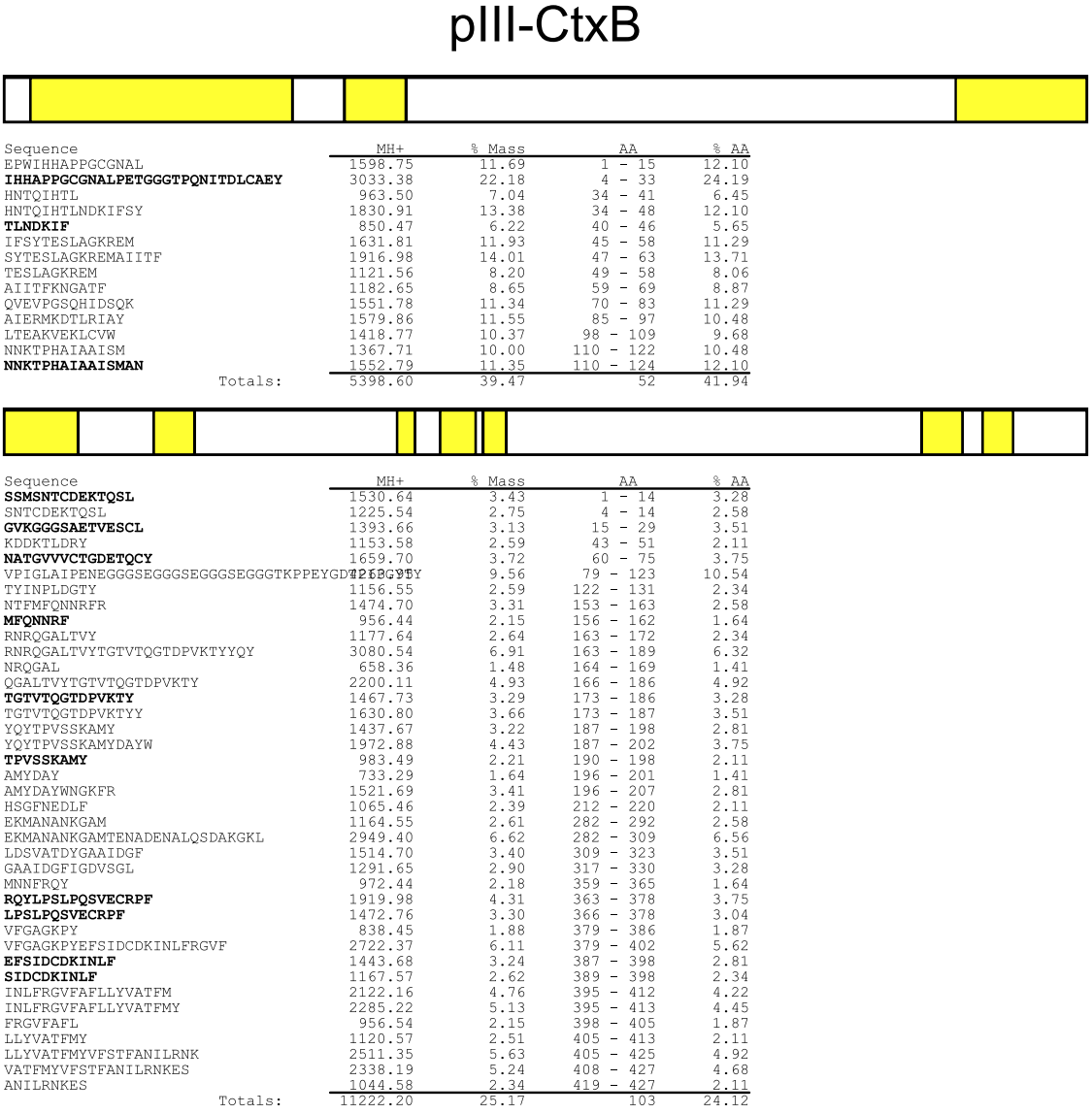
Supplemental Methods

Estimating nearest neighbor packing of GFP on phage surface

Using the crystal structure of the pVIII capsid protein (1IFJ) [121], a model viral capsid was constructed with five fold symmetry serving as a model of the phage surface. A crystal structure of GFP (1GFL) [122] was oriented such that its C-terminus was adjacent to the N-terminus of pVIII. By analyzing this image, it was determined that one GFP molecule blocked the N-termini of the six pVIII proteins surrounding the GFP-pVIII fusion meaning at most one out of seven pVIII proteins can be labeled with a GFP. From this, it was calculated that a single virion with 2700 pVIII proteins would have at most 385 GFP molecules. The visualizations were performed using WinCoot [123].

Table A.1: **Oligonucleotides for N-terminal phage engineering**

Name	**Sequence (5'-3')**
G5pIIIC	GTACCTTTCTATTCTCACTCTGGTGGAGGCGGTGGATC
G5pIIIINC	GGCCGATCCACCGCCTCCACCAGAGTGAGAATAGAAAG
A2G4pVIIIC	GCTGGCGGGGGAGGG
A2G4pVIIINC	GATCCCCTCCCCCGCCAGCTGCA
G5HApIXC	CGGCCATGGCGGGCGGAGGTGGAGGCTACCCATACGATG TTCCAGATTACGCTCAGGG
G5HApIXNC	TGAGCGTAATCTGGAACATCGTATGGGTAGCCTCCACCTC CGCCCGCCATGGCCGGCT
AADSPH-pVIII-Top	GTTCCGATGCTGTCTTTCGCTGCTGCAGATTCGCCGCATA CTGAG
AADSPH-pVIII-Bottom	CTCAGTATGCGGCGAATCTGCAGCAGCGAAAGACAGCATC GGAAC



Sequence	MH+	% Mass	AA	% AA
EPWIHHAPPGCGNAL	1598.75	11.69	1 - 15	12.10
IHHAPPGCGNALPETGGGTPQNITDLCAEY	3033.38	22.18	4 - 33	24.19
HNTQIHTL	963.50	7.04	34 - 41	6.45
HNTQIHTLNDKIFSY	1830.91	13.38	34 - 48	12.10
TLNDKIF	850.47	6.22	40 - 46	5.65
IFSYTESLAGKREM	1631.81	11.93	45 - 58	11.29
SYTESLAGKREMAIITF	1916.98	14.01	47 - 63	13.71
TESLAGKREM	1121.56	8.20	49 - 58	8.06
AIITFKNGATF	1182.65	8.65	59 - 69	8.87
QVEVPGSQHIDSQK	1551.78	11.34	70 - 83	11.29
AIERMKDTLRIAY	1579.86	11.55	85 - 97	10.48
LTEAKVEKLCVW	1418.77	10.37	98 - 109	9.68
NNKTPHAIAAISM	1367.71	10.00	110 - 122	10.48
NNKTPHAIAAISMAN	1552.79	11.35	110 - 124	12.10
Totals:	5398.60	39.47	52	41.94

Sequence	MH+	% Mass	AA	% AA
SSMSNTCDEKTQSL	1530.64	3.43	1 - 14	3.28
SNTCDEKTQSL	1225.54	2.75	4 - 14	2.58
GVKGGGSAETVESCL	1393.66	3.13	15 - 29	3.51
KDDKTLDRY	1153.58	2.59	43 - 51	2.11
NATGVVVCTGDETQCY	1659.70	3.72	60 - 75	3.75
VPIGLAIPENEGGGSEGGGSEGGGSEGGGTKPPEYGDTPIGYY	4268.95	9.56	79 - 123	10.54
TYINPLDGTY	1156.55	2.59	122 - 131	2.34
NTFMFQNNRFR	1474.70	3.31	153 - 163	2.58
MFQNNRF	956.44	2.15	156 - 162	1.64
RNRQGALTVY	1177.64	2.64	163 - 172	2.34
RNRQGALTVYTGTVTQGTDPVKTYYQY	3080.54	6.91	163 - 189	6.32
NRQGAL	658.36	1.48	164 - 169	1.41
QGALTVYTGTVTQGTDPVKTY	2200.11	4.93	166 - 186	4.92
TGTVTQGTDPVKTY	1467.73	3.29	173 - 186	3.28
TGTVTQGTDPVKTYY	1630.80	3.66	173 - 187	3.51
YQYTPVSSKAMY	1437.67	3.22	187 - 198	2.81
YQYTPVSSKAMYDAYW	1972.88	4.43	187 - 202	3.75
TPVSSKAMY	983.49	2.21	190 - 198	2.11
AMYDAY	733.29	1.64	196 - 201	1.41
AMYDAYWNGKFR	1521.69	3.41	196 - 207	2.81
HSGFNEDLF	1065.46	2.39	212 - 220	2.11
EKMANANKGAM	1164.55	2.61	282 - 292	2.58
EKMANANKGAMTENADENALQSDAKGKL	2949.40	6.62	282 - 309	6.56
LDSVATDYGAAIDGF	1514.70	3.40	309 - 323	3.51
GAAIDGFIGDVSGL	1291.65	2.90	317 - 330	3.28
MNNFRQY	972.44	2.18	359 - 365	1.64
RQYLPSLPQSVECRPF	1919.98	4.31	363 - 378	3.75
LPSLPQSVECRPF	1472.76	3.30	366 - 378	3.04
VFGAGKPY	838.45	1.88	379 - 386	1.87
VFGAGKPYEFSIDCDKINLFRGVF	2722.37	6.11	379 - 402	5.62
EFSIDCDKINLF	1443.68	3.24	387 - 398	2.81
SIDCDKINLF	1167.57	2.62	389 - 398	2.34
INLFRGVFAFLLYVATFM	2122.16	4.76	395 - 412	4.22
INLFRGVFAFLLYVATFMY	2285.22	5.13	395 - 413	4.45
FRGVFAFL	956.54	2.15	398 - 405	1.87
LLYVATFMY	1120.57	2.51	405 - 413	2.11
LLYVATFMYVFSTFANILRNK	2511.35	5.63	405 - 425	4.92
VATFMYVFSTFANILRNKES	2338.19	5.24	408 - 427	4.68
ANILRNKES	1044.58	2.34	419 - 427	2.11
Totals:	11222.20	25.17	103	24.12

Figure A.4: **Characterization of the pIII-CtxB conjugate by mass spectrometry**. The polypeptide corresponding to pIII-CtxB was excised from the SDS-PAGE gel and digested with chymotrypsin. Under reducing conditions before digestion, the pIII-CtxB separates into two portions, which are shown in the two panels. The resulting peptides were analyzed by liquid chromatography MS/MS. Peptides positively identified by sequence are highlighted in yellow and bold.

Table A.2: **Oligonucleotides for C-terminal phage engineering**

Name	**Sequence (5'-3')**
pIIIBspEIBclITop	CGTTTGCTAACATACTCCGGAATAAGGA
	GTCTTGATCATGCCAGTTCTTTTGG
pIIIBspEIBclIBottom	CCAAAAGAACTGGCATGATCAAGACTCC
	TTATTCCGGAGTATGTTAGCAAACG
pVIAatIITop	AGGCTGCTATTTTCATTTTTGACGTCAA
	ACAAAAAATCGTTTCTTA
pVIAatIIBottom	TAAGAAACGATTTTTTGTTTGACGTCAA
	AAATGAAAATAGCAGCCT
pVIAgeITop	ATATGGCTGTTTATTTTGTAACCGGTAA
	ATTAGGCTCTGGAAAGAC
pVIAgeIBottom	GTCTTTCCAGAGCCTAATTTACCGGTTA
	CAAAATAAACAGCCATAT
pIXSpeITop	TATTTTACCCGTTTAATGGAAACTAGTT
	CATGAAAAAGTCTTTAGTCC
pIXSpeIBottom	GGACTAAAGACTTTTTCATGAACTAGTT
	TCCATTAAACGGGTAAAATA
pIIILPETGGHA-C	CCGGAATAAGGAGTCTCTACCGGAAACA
	GGAGGCTACCCATACGATGTTCCAGATT
	ACGCTT
pIIILPETGGHA-NC	GATCAAGCGTAATCTGGAACATCGTATG
	GGTAGCCTCCTGTTTCCGGTAGAGACTC
	CTTATT
pIII1GLPETGGHA-C	CCGGAATAAGGAGTCTGGAGGTGGAAGT
	CTACCGGAAACAGGAGGCTACCCATACG
	ATGTTCCAGATTACGCTT
pIII1GLPETGGHA-NC	GATCAAGCGTAATCTGGAACATCGTATG
	GGTAGCCTCCTGTTTCCGGTAGACTTCC
	ACCTCCAGACTCCTTATT
pIII3GLPETGGHA-C	CCGGAATAAGGAGTCTGGAGGTGGAAGT
	GGCGGTGGGAGCGGGGGAGGCTCTCTAC
	CGGAAACAGGAGGCTACCCATACGATGT
	TCCAGATTACGCTT
pIII3GLPETGGHA-NC	GATCAAGCGTAATCTGGAACATCGTATG
	GGTAGCCTCCTGTTTCCGGTAGAGAGCC
	TCCCCCGCTCCCACCGCCACTTCCACCT
	CCAGACTCCTTATT
pVILPETGGHA-C	CAAACAAAAAATCGTTTCTTATTTGGATT
	GGGATAAACTACCGGAAACAGGAGGCTA
	CCCATACGACGTTCCAGATTACGCTTAA
	TATGGCTGTTTATTTTGTAA

Continued on next page

Table A.2 – *Continued from previous page*

Name	**Sequence (5'-3')**
pVILPETGGHA-NC	CCGGTTACAAAATAAACAGCCATATTAAG CGTAATCTGGAACGTCGTATGGGTAGCC TCCTGTTTCCGGTAGTTTATCCCAATCC AAATAAGAAACGATTTTTTGTTTGACGT
pVI1GLPETGGHA-C	CAAACAAAAAATCGTTTCTTATTTGGATT GGGATAAAGGAGGTGGAAGTCTACCGGA AACAGGAGGCTACCCATACGACGTTCCA GATTACGCTTAATATGGCTGTTTATTTT GTAA
pVI1GLPETGGHA-NC	CCGGTTACAAAATAAACAGCCATATTAAG CGTAATCTGGAACGTCGTATGGGTAGCC TCCTGTTTCCGGTAGACTTCCACTCCTT TATCCCAATCCAAATAAGAAACGATTTTT TGTTTGACGT
pVI3GLPETGGHA-C	CAAACAAAAAATCGTTTCTTATTTGGATT GGGATAAAGGAGGTGGAAGTGGCGGTGG GAGCGGGGGAGGCTCTCTACCGGAAACA GGAGGCTACCCATACGACGTTCCAGATT ACGCTTAATATGGCTGTTTATTTTGTAA
pVI3GLPETGGHA-NC	CCGGTTACAAAATAAACAGCCATATTAAG CGTAATCTGGAACGTCGTATGGGTAGCC TCCTGTTTCCGGTAGAGAGCCTCCCCCG CTCCCACCGCCACTTCCACCTCCTTTAT CCCAATCCAAATAAGAAACGATTTTTTG TTTGACGT
pIXLPETGGHA-C	CTAGTTCTCTCCCGGAAACAGGTGGATAC CCATACGATGTTCCAGATTACGCTT
pIXLPETGGHA-NC	CATGAAGCGTAATCTGGAACATCGTATGG GTATCCACCTGTTTCCGGGAGAGAA
pIX1GLPETGGHA-C	CTAGTTCTGGAGGTGGAAGTCTCCCGGAA ACAGGTGGATACCCATACGATGTTCCAG ATTACGCTT
pIX1GLPETGGHA-NC	CATGAAGCGTAATCTGGAACATCGTATGG GTATCCACCTGTTTCCGGGAGACTTCCA CCTCCAGAA
pIX3GLPETGGHA-C	CTAGTTCTGGAGGTGGAAGTGGCGGTGGG AGCGGGGGAGGCTCTCTCCCGGAAACAGG TGGATACCCATACGATGTTCCAGATTA CGCTT

Continued on next page

Table A.2 – *Continued from previous page*

Name	**Sequence (5'-3')**
pIX3GLPETGGHA-NC	CATGAAGCGTAATCTGGAACATCGTATGG GTATCCACCTGTTTCCGGGAGAGAGCCTC CCCCGCTCCCACCGCCACTTCCACCTC CAGAA
96gIII	CCCTCATAGTTAGCGTAACG
pVISeq	GTTGCTATTTTGCACCCAGC
pIIILoop-C	GTACCTTTCTATTCTCACTCTGAGCCGT GGATTCATCATGCACCGCCGGGTTGTGG GAATGCTCTTCCTGAGACCGGTGGTTAC CCATACGATGTTCCAGATTACGCTATGA ATGCTCCAAGATCATCGATGAGTAATAC TTGCGATGAAAAAACCCAAAGTCTAGGT GTAAAAGGAGGCGGGTC
pIIILoop-NC	GGCCGACCCGCCTCCTTTTACACCTAGA CTTTGGGTTTTTTCATCGCAAGTATTAC TCATCGATGATCTTGGAGCATTCATAGC GTAATCTGGAACATCGTATGGGTAACCA CCGGTCTCAGGAAGAGCATTCCCACAAC CCGGCGGTGCATGATGAATCCACGGCTC AGAGTGAGAATAGAAAG
pIIILoopXaTop	GTTCCAGATTACGCTATTGAAGGGAGAT CATCGATGAATAC
pIIILoopXaBottom	GTATTCATCGATGATCTCCCTTCAATAG CGTAATCTGGAAC

Appendix B

Constructs and protocols

2.1 Constructs

Table B.1: **Phage vectors**

Construct	**pIII**	**pVIII**	**pIX**	**Vector**
G_5-pIII	G_5	AEGD	-	M13KE
A_2G_4-pVIII	-	AAGGGG	-	M13SK
G_5HA-pIX	SBP	ADSPHTELP	G_5-(HA)-QGGGQGVD	983
A_2-pVIII G_5-pIII	G_5	AADSPHTELP	-	M13SK
pIIIpVIpIX-CCut	SBP	ADSPHTELP	(BAP)-GQGGGQGVD	983
LoopXa-pIII	LoopXa	AEGD	-	M13KE
A_2-pVIII	BAP	AADSPHTELP	-	M13SK

Note: The sequences listed in the table above represent the N-termini of the pIII, pVIII, and pIX proteins. For pIII and pIX, the sequences are fused to the N-terminus of the mature proteins present in M13KE. For pVIII, the sequences listed are fused to the following sequence: DPAKAAFNSLQASATEYIGYAWAMVVVIVGATIGIKLFKKFTSKAS. SBP (SPARC binding peptide)= SPPTGINGGG. HA= YPYDVPDYA. BAP (biotin acceptor peptide)= GLQDIFEAQKIEWHE. LoopXa= EPWIHHAPPGCGNALPETGGYPYDVPDYAIEGRSSMSNTCDEKTQSLGVKGGG.

Table B.2: **Protein expression vectors**.

Protein	**Antibiotic**	**Inducer**
$SrtA_{pyogenes}$	Kanamycin	IPTG
$SrtA_{aureus}$	Ampicillin	IPTG
GFP.LPETG.H_6	Kanamycin	IPTG
GFP.LPETAA.H_6	Kanamycin	IPTG
GFP.HAtag.LPETG.H_6	Kanamycin	IPTG
GFP.HAtag.LPETAA.H_6	Kanamycin	IPTG
Streptavidin.HAtag.LPETGG.H_6	Kanamycin	IPTG
G_3.CtxB	Kanamycin	Arabinose

2.2 Protocols

Amplification Of M13 Phage

Materials

1. ***E. coli***. Strain ER 2738.
2. **LB Broth**. 25g of LB broth powder per 1L of H_2O
3. **1000X tetracycline**. 30mg/mL in methanol
4. **PEG8000/ NaCl**. 150g PEG8000, 110g NaCl in 750mL of H_2O
5. **TBS**. 25mM Tris, 150mM NaCl, pH 7.0-7.4
6. **M13 Bacteriophage**. $\sim 10^{11}$-10^{12} pfu/mL

Protocol

Day 0

Note: Small scale amplification amounts are indicated in parentheses.

1. Prepare overnight culture (OC): 20mL LB, 20μL 1000X tetracycline, 1 colony of *E. coli* (ER 2738) in sterilized 250 mL flask. Shake at 37°C overnight (12-16 hrs).
2. Prepare 800mL (25mL) of LB broth in a 2L (250mL) flask.

Day 1

1. Dilute 8mL (200μL) OC with 800μL (20μL) of 1000X tetracycline in each 2L (250mL) flask with prepared LB broth.
2. Infect bacteria with 70μL (15μL) of 10^{11}-10^{12} pfu/mL phage solution for the 2L (250mL) flask.
3. Place flasks in 37°C incubator for 30min. After incubation, shake at 37°C for >12hrs.
4. Pour contents of flask into sterilized centrifuge tubes. Balance tubes with LB, and spin in table top centrifuge at 12000g for 15min to pellet out bacteria.
5. Mix supernatant with 160mL (5mL) PEG/NaCl. Invert 4-6 times, and store at 4°C overnight.

Day 2

1. Spin phage-PEG/NaCl mixture at 13500g for 20min to pellet out phage.
2. Carefully pour out and discard supernatant. Resuspend phage precipitates in 15mL (1mL) of TBS.
3. Transfer resuspended phage from into 50mL (1.5mL) centrifuge tubes. Spin at 12000g at 4°C for 15min to pellet out bacteria.
4. Pour supernatant into new sterilized centrifuge tubes, taking care not to disrupt bacterial pellet. 5mL of PEG was added to supernatant. Invert 4-6 times, and store at 4°C overnight, or on ice for 2 hrs. (*Note: The small scale amplification is complete at this point and can be stored at* $4°C$).

Day 3

1. Spin phage-PEG/NaCl solution at 13500g for 20min to pellet out phage.
2. Pour out and discard supernatant. Resuspend phage precipitates in 1mL of TBS.
3. Transfer resuspended phage from the centrifuge tube to a 1.5mL tube.
4. Centrifuge phage solution in microcentrifuge at 10000g for 15min.
5. Transfer supernatant into another clean 1.5mL tube. This final solution should have a concentration of 10^{13}-10^{14} pfu/mL.

Purification of phage sortase reaction

Materials

1. **Sortase reaction**. 30μL in TBS.
2. **PEG8000/NaCl**. 150g PEG8000, 110g NaCl in 750mL water
3. **TBS**. 25mM Tris, 150mM NaCl, pH 7.0-7.4

Protocol

1. Dilute reaction to 1mL in TBS in a 1.5mL eppendorf tube. For reactions greater than 30μL divide reaction into multiple tubes.
2. Move diluted mixture to another eppendorf containing 250μL of PEG8000/NaCl solution. Mix well.
3. Incubate mixture on ice for at least 3hrs.
4. Centrifuge sample at 4°C for 17000g for 20min. Remove supernatant and spin for an additional minute to remove excess PEG and salt.
5. After removing the supernatant, resuspend the pellet in desired buffer.

Creating DNA peptide conjugates

Materials

1. **DNA-thiol**. 100μM in PBS.
2. **Maleimide-peptide**. 1mM in PBS.
3. **Dithiothreitol (DTT)**. 1M in water
4. **NAP5 Column**. GE Healthcare
5. **TBS**. 25mM Tris, 150mM NaCl, pH 7.0-7.4

Protocol

Day 0

1. 50μL of 100μM DNA-thiol is mixed with 5μL of 1M DTT. The mixture is incubated at 37°C overnight (>12hrs).

Day 1

1. Equilibrate NAP5 column with 10mL of water. Dilute DTT and DNA mixture to 500μL in water. Add to column and elute with 1mL of water.
2. Dry 1mL to a pellet by spin vacuum. Resuspend pellet in 50μL of PBS. Measure concentration of DNA solution using absorbance at 260nm and extinction coefficient of DNA oligonucleotide.
3. Add maleimide-peptide such that a 2:1 molar ratio of peptide to DNA. Incubate the mixture at 37°C for 5hrs.
4. After 5hrs add 5μL of 1M DTT and incubate at 37°C for 15-30min.
5. Equilibrate a NAP5 column with 10mL of water. Dilute the DNA, peptide, and DTT mixture in water to 500μL and add to the column. Elute with 1mL of water.
6. Dry the elution by spin vacuum and resuspend in 50μL of TBS. Store at 4°C.

Table B.3: **Antibodies for immunoblotting**.

Antibody	**Blocking Buffer**	**Concentration**	**Secondary**
anti-pIII	Milk	1:8000	anti-mouse-HRP (1:8000)
anti-HA-HRP	Milk	1:6000	
streptavidin-HRP	BSA	1:15000	
anti-GFP-HRP	Milk	1:10000	
anti-CtxB	Milk	1:6000	anti-rabbit-HRP (1:6000)

Bibliography

[1] J. D. Currey. Mechanical properties of mother of pearl in tension. *Proc. R. Soc. London Ser. B.*, 196:443–463, 1977.

[2] G. Mayer. Rigid biological systems as models for synthetic composites. *Science*, 310:1144–1147, 2005.

[3] L. J. Bonderer, A. R. Studart, and L. J. Gauckler. Bioinspired design and assembly of platelet reinforced polymer films. *Science*, 319:1069–1073, 2008.

[4] S. Zhang. Fabrication of novel biomaterials through molecular self-assembly. *Nat. Biotechnol.*, 21:1171–1178, 2003.

[5] C. Mao, C. E. Flynn, A. Hayhurst, R. Sweeney, J. Qi, G. Georgiou, B. Iverson, and A. M. Belcher. Viral assembly of oriented quantum dot nanowires. *Proc. Natl. Acad. Sci. U.S.A.*, 100:6946–6951, 2003.

[6] C. Mao, D. J. Solis, B. D. Reiss, S. T. Kottmann, R. Y. Sweeney, A. Hayhurst, G. Georgiou, B. Iversion, and Belcher A. M. Virus-based toolkit for the directed synthesis of magnetic and semiconducting nanowires. *Science*, 303:213–217, 2004.

[7] K. T. Nam, B. R. Peelle, S.-W. Lee, and A. M. Belcher. Genetically driven assembly of nanorings based on the M13 virus. *Nano Lett.*, 4:23–27, 2004.

[8] P. J. Yoo, K. T. Nam, J. Qi, Lee. S.-K., J. Park, A. M Belcher, and P. T. Hammond. Spontaneous assembly of viruses on multilayered polymer surfaces. *Nat. Mater.*, 5:234–240, 2006.

[9] E. Winfree, F. Liu, L. A. Wenzler, and N. C. Seeman. Design and self-assembly of two-dimensional DNA crystals. *Nature*, 394:539–544, 1998.

[10] S. M. Douglas, H. Dietz, T. Liedl, B. Hogberg, F. Graf, and W. M. Shih. Self-assembly of DNA into nanoscale three-dimensional shapes. *Nature*, 459:414–418, 2009.

[11] N. C. Seeman. An overview of structural DNA nanotechnology. *Mol. Biotechnol.*, 37:246–257, 2007.

[12] S. Modi, D. Bhatia, F. C. Simmel, and Y. Krishnan. Structural DNA nanotechnology: from bases to bricks, from structures to function. *J. Phys. Chem. Lett.*, 1:1994–2005, 2010.

[13] P. W. K. Rothemund. Folding DNA to create nanoscale shapes and patterns. *Nature*, 440:297–302, 2006.

[14] S. Modi, M. G. Swetha, D. Goswami, G. D. Gupta, S. Mayor, and Y. Krishnan. A DNA nanomachine that maps spatial and temporal pH changes inside living cells. *Nat. Nanotechnol.*, 4:325–330, 2009.

[15] S. M. Douglas, I. Bachelet, and G. M. Church. A logic-gated nanorobot for targeted transport of molecular payloads. *Science*, 335:831–834, 2012.

[16] H. O. Smith and K. W. Wilcox. A restriction enzyme from *Hemophilus influenzae*: I. purification and general properties. *J. Mol. Biol.*, 51:379–391, 1970.

[17] R. J. Roberts. How restriction enzymes become the workhorses of molecular biology. *Proc. Natl. Acad. Sci. U.S.A.*, 102:5905–5908, 2005.

[18] D. Beckett, E. Kovaleva, and P. J. Schatz. A minimal peptide substrate in biotin holoenzyme synthetase-catalyzed biotinylation. *Protein Sci.*, 8:921–929, 1999.

[19] A. Chapman-Smith and J. E. Cronan, Jr. The enzymatic biotinylation of proteins: a post-translational modification of exception specificity. *Trends Biochem. Sci.*, 24:359–363, 1999.

[20] K. J. Roux, D. I. Kim, M. Raida, and B. Burke. A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. *J. Cell Biol.*, 196:801–810, 2012.

[21] D. A. Marvin. Filamentous phage structure, infection and assembly. *Curr. Opin. Struct. Biol.*, 8:150–158, 1998.

[22] G. P. Smith. Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science*, 228:1315–1317, 1985.

[23] V. A. Petrenko, G. P. Smith, X. Gong, and T. Quinn. A library of organic landscapes on filamentous phage. *Protein Eng.*, 9:797–801, 1996.

[24] C. Krebber, S. Spada, D. Desplancq, and A. Pluckthun. Co-selection of cognate antibody-antigen pairs by selectively-infective phages. *FEBS Lett.*, 377:227–231, 1995.

[25] J. J. Devlin, L. C. Panganiban, and P. E. Devlin. Random peptide libraries: a source of specific protein binding molecules. *Science*, 249:404–406, 1990.

[26] L. Zang, L. Shi, J. Guo, Q. Pan, W. Wu, X. Pan, and J. Wang. Screening and identification of a peptide specifically targeted to NCI-H1299 from a phage display peptide library. *Cancer Lett.*, 281:64–70, 2009.

[27] S. R. Whaley, D. S. English, E. L. Hu, P. F. Barbara, and A. M. Belcher. Selection of peptides with semiconductor binding specificity for directed nanocrystal assembly. *Nature*, 405:665–668, 2000.

[28] Y. Huang, C. Y. Chiang, S.-K. Lee, Y. Gao, E. L. Hu, J. De Yoreo, and A. M. Belcher. Programmable assembly of nanoarchitectures using genetically engineered viruses. *Nano Lett.*, 5:1429–1434, 2005.

[29] K. A. Kelly, P. Waterman, and R. Weissleder. *In vivo* imaging of molecularly targeted phage. *Neoplasia*, 8:1011–1018, 2006.

[30] H. Yi, D. Ghosh, M.-H. Ham, J. Qi, P. W. Barone, M. S. Strano, and A. M. Belcher. M13 phage-functionalized single-walled carbon nanotubes as nanoprobes for second near-infrared window fluorescence imaging of targeted tumors. *Nano Lett.*, 12:1176–1183, 2012.

[31] D. Larocca, A. Witte, W. Johnson, G. F. Pierce, and A. Baird. Targeting bacteriophage to mammalian cell surface receptors for gene delivery. *Hum. Gene Ther.*, 9:2393–2399, 1998.

[32] H. Bar, I. Yacoby, and I. Benhar. Killing cancer cells by targeted drug-carrying phage nanomedicines. *BMC Biotechnol.*, 8:37, 2008.

[33] J. Greenwood, A. E. Willis, and R. N. Perham. Multiple display of foreign peptides on a filamentous bacteriophage. Peptides from Plasmodium falciparum circumsporozoite protein as antigens. *J. Mol. Biol.*, 220:821–827, 1991.

[34] G. Iannolo, O. Minenkova, R. Petruzzelli, and G. Cesareni. Modifying filamentous phage capsid: limits in the size of the major capsid protein. *J. Mol. Biol.*, 248:835–844, 1995.

[35] K. T. Nam, D. W. Kim, P. J. Yoo, C. Y. Chiang, N. Meethong, P. T. Hammond, Y. M. Chang, and A. M. Belcher. Virus enabled synthesis and assembly of nanowires for lithium ion battery electrodes. *Science*, 312:885–888, 2006.

[36] Y. J. Lee, H. Yi, W. J. Kim, K. Kang, D. S. Yun, M. S. Strano, G. Ceder, and A. M. Belcher. Fabricating genetically engineered high-power lithium ion batteries using multiple virus genes. *Science*, 324:1051–1055, 2009.

[37] K. T. Nam, Y. J. Lee, E. M. Krauland, S. T. Kottmann, and A. M. Belcher. Peptide-mediated reduction of silver ions on engineered biological scaffolds. *ACS Nano*, 2:1480–1486, 2008.

[38] X. Dang, H. Yi, M.-H. Ham, J. Qi, D. S. Yun, R. Ladewski, M. S. Strano, P. T. Hammond, and A. M. Belcher. Virus-templated self-assembled single-walled carbon nanotubes for highly efficient electron collection in photovoltaic devices. *Nat. Nanotechnol.*, 6:377–384, 2011.

[39] Y. S. Nam, A. P. Magyar, D. Lee, J.-W. Kim, D. S. Yun, H. Park, T. S. Pollom, Jr., D. A. Weitz, and A. M. Belcher. Biologically templated photocatalytic nanostructures for sustained light-drive water oxidation. *Nat. Nanotechnol.*, 5:340–344, 2010.

[40] S. Ng, M. R. Jafari, and R. Derda. Bacteriophages and viruses as a support for organic synthesis and combinatorial chemistry. *ACS Chem. Biol.*, 7:123–138, 2011.

[41] K. R. Love, J. G. Swoboda, C. J. Noren, and S. Walker. Enabling glycosyltransferase evolution: a facile substrate-attachment strategy for phage-display enzyme evolution. *Chembiochem*, 7:753–756, 2006.

[42] K. L. Kiick, E. Saxon, D. A. Tirrell, and C. R. Bertozzi. Incorporation of azides into recombinant proteins for chemoselective modification by the Staudinger ligation. *Proc. Natl. Acad. Sci. U.S.A.*, 99:19–24, 2002.

[43] S. K. Mazmanian, G. Liu, H. Ton-That, and O. Schneewind. *Staphylococcus aureus* sortase, an enzyme that anchors surface proteins to the cell wall. *Science*, 285:760–763, 1999.

[44] H. Ton-That, G. Liu, S. K. Mazmanian, K. F. Faull, and O. Schneewind. Purification and characterization of sortase, the transpeptidase that cleaves surface proteins of *Staphylococcus aureus* at the LPXTG motif. *Proc. Natl. Acad. Sci. USA*, 96:12424–12429, 1999.

[45] P. R. Race, M. L. Bentley, J. A. Melvin, A. Crow, R. K. Hughes, W. D. Smith, R. B. Sessions, M. A. Kehoe, D. G. McCafferty, and M. J. Banfield. Crystal structure of *Streptococcus pyogenes* sortase A: implications for sortase mechanism. *J. Biol. Chem.*, 284:6924–6933, 2009.

[46] M. W. Popp, J. M. Anos, G. M. Grotenbreg, E. Spooner, and H. L. Ploegh. Sortagging: a versatile method for protein labeling. *Nat. Chem. Biol.*, 3:707–708, 2007.

[47] J. M. Antos, G.-L. Chew, C. P. Guimaraes, N. C. Yoder, G. M. Grotenbreg, M. W. Popp, and H. L. Ploegh. Site-specific N- and C-terminal labeling of a single polypeptide using sortases of different specificity. *J. Am. Chem. Soc.*, 131:10800–10801, 2009.

[48] C. P. Guimaraes, J. E. Carette, M. Varadarajan, J. M. Antos, M. W. Popp, E. Spooner, T. R. Brummelkamp, and H. L. Ploegh. Identification of host cell factors required for intoxication through use of modified cholera toxin. *J. Cell Biol.*, 195:751–764, 2011.

[49] M. W. Popp, S. K. Dougan, T.-Y. Chuang, E. Spooner, and H. L. Ploegh. Sortase-catalyzed transformations that improve the properties of cytokines. *Proc. Natl. Acad. Sci. U.S.A.*, 108:3169–3174, 2011.

[50] J. M. Antos, G. M. Miller, G. M. Grotenbreg, and H. L. Ploegh. Lipid modification of proteins through sortase-catalyzed transpeptidation. *J. Am. Chem. Soc.*, 130:16338–16343, 2008.

[51] M. W. Popp and H. L. Ploegh. Making and breaking peptide bonds: protein engineering using sortase. *Angew. Chem. Int. Ed. Engl.*, 50:5024–5032, 2011.

[52] G. T. Hess, J. J. Cragnolini, M. W. Popp, M. A. Allen, S. K. Dougan, E. Spooner, H. L. Ploegh, A. M. Belcher, and C. P. Guimaraes. An M13 bacteriophage display framework that allows sortase-mediated modification of surface-accessible phage proteins. *Bioconj. Chem.*, in press.

[53] S. S. Sidhu. Engineering M13 for phage display. *Biomol. Eng.*, 18:57–63, 2001.

[54] T. Bratkovic. Progress in phage display: evolution of the technique and its application. *Cell. Mol. Life Sci.*, 67:749–767, 2010.

[55] J. B. Burritt, M. T. Quinn, M. A. Jutila, C. W. Bond, and A. J. Jesaitis. Topological mapping of neutrophil cytochrome *b* epitopes with phage-display libraries. *J. Biol. Chem.*, 270:16974–16980, 1995.

[56] M. A. Barry, W. J. Dower, and S. A. Johnston. Toward cell-targeting gene therapy vectors: selection of cell-binding peptides from random peptide-presenting phage libraries. *Nat. Med.*, 2:299–305, 1996.

[57] D. L. Jaye, F. S. Nolte, L. Mazzucchelli, C. Geigerman, A. Akyildiz, and C. A. Parkos. Use of real-time polymerase chain reaction to identify cell- and tissue-type-selective peptides by phage display. *Am. J. Pathol.*, 162:1419–1429, 2003.

[58] L. Mazzucchelli, J. B. Burritt, A. J. Jesaitis, A. Nusrat, T. W. Liang, A. T. Gewirtz, F. J. Schnell, and C. A. Parkos. Cell-specific peptide binding by human neutrophils. *Blood*, 93:1738–1748, 1999.

[59] A. K. Udit, W. Hollingsworth, and K. Choi. Metal- and metallocycle-binding sites engineered into polyvalent virus-like scaffolds. *Bioconjug. Chem.*, 21:399–404, 2010.

[60] E. Kaltgrad, M. K. O'Reilly, L. Liao, S. Han, J. C. Paulson, and M. G. Finn. On-virus construction of polyvalent glycan ligands for cell-surface receptors. *J. Am. Chem. Soc.*, 130:4578–4579, 2008.

[61] E. Bianchi, A. Folgori, A. Wallace, M. Nicotra, S. Acali, A. Phalipon, G. Barbato, R. Bazzo, R. Cortese, F. Felici, and A. Pessi. A conformationally homogenous ombinatorial peptide library. *J. Mol. Biol.*, 247:154–160, 1995.

[62] D. R. Corey, A. K. Shiau, Q. Yang, B. A. Janowski, and C. S. Craik. Trypsin display on the surface of bacteriophage. *Gene*, 128:129–134, 1993.

[63] A. S. Kang, C. F. Barbas, K. D. Janda, S. J. Benkovic, and R. A. Lerner. Linkage of recognition and reputation functions by assembling combinatorial antibody Fab libraries along phage surfaces. *Proc. Natl. Acad. Sci. U.S.A.*, 88:4363–4366, 1991.

[64] P. Malik, T. D. Terry, L. R. Gowda, A. Langara, S. A. Petukhov, M. F. Symmons, L. C. Welsh, D. A. Marvin, and R. N. Perham. Role of capsid structure and membrane protein processing in determining the size and copy number of

peptides displayed on the major coat protein of filamentous bacteriophage. *J. Mol. Biol.*, 260:9–21, 1996.

[65] W. Markland, B. L. Roberts, M. J. Saxena, S. K. Guterman, and R. C. Ladner. Design, construction and function of a multicopy display vector using fusions to the major coat protein of bacteriophage M13. *Gene*, 109:13–19, 1991.

[66] S. Bass, R. Greene, and J. A. Wells. Hormone phage: an enrichment method for variant proteins with altered binding properties. *Proteins*, 296:309–314, 1990.

[67] S. S. Sidhu, G. A. Weiss, and J. A. Wells. High copy display of large proteins on phage for functional selections. *J. Mol. Biol.*, 296:487–495, 2000.

[68] T. Kretzschmar and M. Geiser. Evaluation of antibodies fused to minor coat protein III and major coat protein VIII of bacteriophage M13. *Gene*, 155:61–65, 1995.

[69] C. Gao, S. Mao, C.-H. L. Lo, P. Wirsching, R. A. Lerner, and K. D. Janda. Making artificial antibodies: a format for phage display of combinatorial heterodimeric arrays. *Proc. Natl. Acad. Sci. USA*, 96:6025–6030, 1999.

[70] L. S. Jespers, J. H. Messens, A. De Keyser, D. Eeckhout, I. Van den Brande, Y. G. Gansemans, M. J. Lauwereys, G. P. Vlasuk, and P. E. Stanssens. Surface expression and ligand-based selection of cDNAs fused to filamentous phage gene VI. *Biotechnology (NY)*, 13:378–382, 1995.

[71] Y. Georgieva and Z. Konthur. Design and screening of M13 phage display cDNA libraries. *Molecules*, 16:1667–1681, 2011.

[72] S. Zozulya, M. Lioubin, R. J. Hill, C. Abram, and M. L. Gishizky. Mapping signal transduction pathways by phage display. *Nat. Biotechnol.*, 17:1193–1198, 1999.

[73] H. Ton-That, S. K. Mazmanian, K. F. Faull, and O. Schneewind. Anchoring of surface proteins to the cell wall of *Staphylococcus aureus*. sortase catalyzed *in vitro* transpeptidation reaction using LPXTG peptide and NH_2-Gly_3 substrates. *J. Biol. Chem.*, 275:9876–9881, 2000.

[74] J. Lubkowski, F. Hennecke, A. Pluckthun, and A. Wlodawer. The structural basis of phage display elucidated by the crystal structure of the N-terminal domains of g3p. *Nat. Struct. Biol.*, 5:140–147, 1998.

[75] L. Makowski. Terminating a macromolecular helix: structural model for the minor proteins of bacteriophage M13. *J. Mol. Biol.*, 228:885–892, 1992.

[76] C. F. Barbas, D. R. Burton, J. K. Scott, and G. J. Silverman. *Phage Display: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 2001.

[77] D. O'Connell, B. Becerril, A. Roy-Burman, M. Daws, and J. D. Marks. Phage versus phagemid libraries for generation of human monoclonal antibodies. *J. Mol. Biol.*, 321:49–56, 2002.

[78] S. Rondot, J. Koch, F. Breitling, and S. Dubel. A helper phage to improve single-chain antibody presentation in phage display. *Nat. Biotechnol.*, 19:75–78, 2001.

[79] K. A. Kelly, S. R. Setlur, R. Ross, R. Anbazhagan, P. Waterman, M. A. Rubin, and R. Weissleder. Detection of early prostate cancer using a hepsin-targeted imaging agent. *Cancer Res.*, 68:2286–2291, 2008.

[80] D. L. Jaye, C. M. Geigerman, R. E. Fuller, A. Akyildiz, and C. A. Parkos. Direct fluorochrome labeling of phage display library clones for studying binding specificites: applications in flow cytometry and fluorescence microscopy. *J. Immunol. Methods*, 295:119–127, 2004.

[81] A. Kirchhofer, J. Helma, K. Schmidthals, C. Frauer, S. Cui, A. Karcher, M. Pellis, S. Muyldermans, C. S. Casas-Delucchi, M. C. Cardoso, H. Leonhardt, K. P. Hopfner, and U. Rothbauer. Modulation of protein properties in living cells using nanobodies. *Nat. Struct. Mol. Biol.*, 17:133–138, 2010.

[82] P. J. Schatz. Use of peptide libraries to map the substrate specificity of a peptide-modifying enzyme: a 13 residue consensus peptide specifies biotinylation in *Escherichia coli*. *Biotechnology (NY)*, 11:1138–1143, 1993.

[83] T. Matsumoto, S. Sawamoto, T. Sakamoto, T. Tanaka, H. Fukuda, and A. Kondo. Site-specific tetrameric streptavidin-protein conjugation using sortase A. *J. Biotechnol.*, 152:37–42, 2011.

[84] Y. S. Nam, T. Shin, H. Park, A. P. Magyar, K. Choi, G. Fantner, K. A. Nelson, and A. M. Belcher. Virus-templated assembly of porphyrins into light-harvesting nanoantennae. *J. Am. Chem. Soc.*, 132:1462–1463, 2010.

[85] M. W. Popp, J. M. Antos, and H. L. Ploegh. Site-specific protein labeling via sortase-mediated transpeptidation. *Curr. Protoc. Protein Sci.*, Chapter 15:Unit 15–3, 2009.

[86] C. V. Lee, S. S. Sidhu, and G. Fuh. Bivalent antibody phage display mimics natural immunoglobulin. *J. Immunol. Methods*, 284:119–132, 2004.

[87] M. Howarth, D. J. Chinnapen, K. Gerrow, P. C. Dorrestein, M. R. Grandy, N. L. Kelleher, A. El-Husseini, and A. Y. Ting. A monovalent streptavidin with a single femtomolar biotin binding site. *Nat. Methods*, 3:267–273, 2006.

[88] G. Fuh and S. S. Sidhu. Efficient phage display of polypeptides fused to the carboxy-terminus of the M13 gene-3 minor coat protein. *FEBS Lett.*, 480:231–234, 2000.

[89] G. Fuh, M. T. Pisabarro, Y. Li, C. Quan, L. A. Lasky, and S. S. Sidhu. Analysis of PDZ domain-ligand interactions using carboyxl-terminal phage display. *J. Biol. Chem.*, 275:21486–21491, 2000.

[90] S. Munro and H. R. B. Pelham. A C-terminal signal prevents secretion of luminal ER proteins. *Cell*, 48:899–907, 1987.

[91] E. Ruoslahti. RGD and other recognition sequences for integrins. *Annu. Rev. Cell Dev. Biol.*, 12:697–715, 1996.

[92] R.-G. Zhang, M. L. Westbrook, E. M. Westbrook, D. L. Scott, Z. Otwinowski, P. R. Maulik, R. A. Reed, and G. G. Shipley. The 2.4 Å crystal structure of cholera toxin B subunit pentamer: choleragenoid. *J. Mol. Biol.*, 251:550–562, 1995.

[93] S. G. Gould, G. A. Keller, and S. Subramani. Identification of a peroxisomal targeting signal at the carboxy terminus of firefly luciferase. *J. Cell Biol.*, 105:2923–2931, 1987.

[94] A. Kurakin, A. Swistowski, S. C. Wu, and D. E. Bredesen. The PDZ domain as a complex adaptive system. *PLoS One*, 2:e953, 2007.

[95] S. C. Sharma, A. Memic, C. N. Rupasinghe, A.-C. E. Duc, and M. R. Spaller. T7 phage display as a method of peptide ligand discovery for PDZ domain proteins. *Biopolymers*, 92:183–193, 2009.

[96] V. V. Ivanenkov, F. Felici, and A. G. Menon. Targeted delivery of multivalent phage display vectors into mammalian cells. *Biochim. Biophys. Acta*, 1448:463–472, 1999.

[97] V. V. Ivanenkov, F. Felici, and A. G. Menon. Update and intracellular fate of phage display vectors in mammalian cells. *Biochim. Biophys. Acta*, 1448:450–462, 1999.

[98] S.-W. Lee, C. Mao, C. E. Flynn, and A. M. Belcher. Ordering of quantum dots using genetically engineered viruses. *Science*, 296:892–895, 2002.

[99] R. Y. Sweeney, E. Y. Park, B. L. Iverson, and G. Georgiou. Assembly of multimeric phage nanostructures through leucine zipper interactions. *Biotechnol. Bioeng.*, 95:539–545, 2006.

[100] H. G. Craighead. Nanoelectromechanical systems. *Science*, 290:1532–1535, 2000.

[101] S.-S. Lee, L.-S. Huang, C.-J. Kim, and M. C. Wu. Free-space fiber-optic switches based on mems vertical torsion mirrors. *J. Lightw. Technol.*, 17:7–13, 1999.

[102] M. C. Y. Huang, Y. Zhou, and C. J. Chang-Hasnain. A nanoelectromechanical tunable laser. *Nat. Photonics*, 2:180–184, 2008.

[103] W. Srituravanich, L. Pan, Y. Wang, C. Sun, D. B. Bogy, and X. Zhang. Flying plasmonic lens in the near field for high-speed nanolithography. *Nat. Nanotechnol.*, 3:733–737, 2008.

[104] M. Rycenga, C. M. Cobley, J. Zeng, W. Li, C. H. Moran, Q. Zhang, D. Qin, and Y. Xia. Controlling the synthesis and assembly of silver nanostructures for plasmonic aaplications. *Chem. Rev.*, 111:3669–3712, 2011.

[105] A. Pimpin and W. Srituravanich. Review on micro- and nanolithography techniques and their applications. *Eng. J.*, 16:37–56, 2012.

[106] J. Yao, A.-P. Le, S. K. Gray, J. S. Moore, J. A. Rogers, and R. G. Nuzzo. Functional nanostructured plasmonic materials. *Adv. Mater.*, 22:1102–1110, 2010.

[107] D. Nykypanchuk, M. M. Maye, D. van der Lelie, and O. Gang. DNA-guided crystallization of colloidal nanoparticles. *Nature*, 451:549–552, 2008.

[108] D. Sun and O. Gang. Binary heterogeneous superlattices assembled from quantum dots and gold nanoparticles. *J. Am. Chem. Soc.*, 133:5252–5254, 2011.

[109] P. Cigler, A. K. R. Lytton-Jean, D. G. Anderson, M. G. Finn, and S. Y. Park. DNA-controlled assembly of a NaTl lattice structure from gold nanoparticles and protein nanoparticles. *Nat. Mater.*, 9:918–922, 2010.

[110] S. Y. Park, A. K. R. Lytton-Jean, B. Lee, S. Weigand, G. C. Schatz, and C. A. Mirkin. DNA-programmable nanoparticle crystallization. *Nature*, 451:553–556, 2008.

[111] R. J. Macfarlane, M. R. Jones, A. J. Senesi, K. L. Young, B. Lee, J. Wu, and C. A. Mirkin. Establishing the design rules for DNA-mediated programmable colloidal crystallization. *Angew. Chem. Int. Ed. Engl.*, 49:4589–4592, 2010.

[112] C. Y. Chiang, C. M. Mello, J. Gu, E. C. C. M. Silva, K. J. Van Vliet, and A. M. Belcher. Weaving genetically engineered functionality into mechanically robust virus fibers. *Adv. Mater.*, 19:826–832, 2007.

[113] H. A. Atwater and A. Polman. Plasmonics for improved photovoltaic devices. *Nat. Mater.*, 9:205–213, 2010.

[114] S. Scarano, M. Mascini, A. P. F. Turner, and M. Minunni. Surface plasmon resonance imaging for affinity-based biosensors. *Biosens. Bioelectron.*, 25:957–966, 2010.

[115] A. V. Kabashin, P. Evans, S. Pastkovsky, W. Hendren, G. A. Wurtz, R. Atkinson, R. Pollard, V. A. Podolskiy, and A. V. Zayats. Plasmonic nanorod metamaterials for biosensing. *Nat. Mater.*, 8:867–871, 2009.

[116] J. Mertz. Radiative absorption, fluorescence and scattering of a classical dipole near a lossless interface: a unified description. *J. Opt. Soc. Am. B*, 17:1906–1913, 2000.

[117] J. B. Ravnsbaek, M. F. Jacobsen, C. B. Rosen, N. V. Voight, and K. V. Gothelf. DNA-programmed Glaser-Eglinton reactions for the synthesis of conjugated molecular wires. *Angew. Chem. Intl. Ed. Engl*, 50:10851–10854, 2011.

[118] P. Yin, H. M. T. Choi, C. R. Calvert, and N. A. Pierce. Programming biomolecular self-assembly pathways. *Nature*, 451:318–322, 2008.

[119] A. S. Khalil, J. M. Ferrer, R. R. Brau, S. T. Kottmann, C. J. Noren, M. J. Lang, and A. M. Belcher. Single M13 bacteriophage tethering and stretching. *Proc. Natl. Acad. Sci. U.S.A.*, 104:4892–4897, 2007.

[120] Y. Lu, B. R. Goldsmith, N. J. Kybert, and A. T. C. Johnson. DNA-decorated graphene chemical sensors. *Appl. Phys. Lett.*, 97:083107, 2010.

[121] D. A. Marvin, R. D. Hale, C. Nave, and M. Helmer-Citterich. Molecular models and structural comparisons of native and mutant class I filamentous bacteriophages Ff (fd, f1, M13), If1 and IKe. *J. Mol. Biol.*, 235:260–286, 1994.

[122] F. Yang, L. G. Moss, and G. N. Phillips, Jr. The molecular structure of green fluorescent protein. *Nat. Biotechnol.*, 14:1246–1251, 1996.

[123] P. Emsley, B. Lohkamp, W. G. Scott, and K. Cowtan. Features and development of coot. *Acta Crystallogr. D Biol. Crystallogr.*, 66:486–501, 2010.